



US005846806A

United States Patent [19]

Jones et al.

[11] **Patent Number:** **5,846,806**

[45] **Date of Patent:** **Dec. 8, 1998**

[54] **IDENTIFICATION OF A HUMAN CYTOMEGALOVIRUS GENE REGION INVOLVED IN DOWN-REGULATION OF MHC CLASS I HEAVY CHAIN EXPRESSION**

[75] Inventors: **Thomas R. Jones**, Nyack, N.Y.; **Ann E. Campbell**, Norfolk, Va.

[73] Assignees: **American Cyanamid Company**, Madison, N.J.; **Eastern Virginia Medical School**, Norfolk, Va.

[21] Appl. No.: **282,696**

[22] Filed: **Jul. 29, 1994**

[51] **Int. Cl.⁶** **C12N 7/04**

[52] **U.S. Cl.** **435/236**

[58] **Field of Search** 435/172.3, 235.1, 435/236, 320.1; 424/230.1

[56] **References Cited**

FOREIGN PATENT DOCUMENTS

0521427 1/1993 European Pat. Off. .
WO89/10966 11/1989 WIPO .

OTHER PUBLICATIONS

Chee M.S. et al., *Current Topics in Microbiology and Immunology*, vol. 154, 1990, pp. 126–169.
Gilbert, M.J. et al., *J. Virology*, vol. 67, No. 6, 1993, pp. 3461–3469.
Beersma, M.F.C. et al., *J. Immunology*, vol. 151, No. 9, 1993, pp. 4455–4464.
Jones et al., T.R. et al., *J. Virology*, vol. 69, No. 8, 1995, pp. 4830–4841.

Colberg-Poley, A. M. et al., *J. Virology*, vol. 66, No. 1, 1992, pp. 95–105.

Jones et al., *J. of Virology*, Apr. 1992, vol. 66(4): pp. 2541–2546.

Kollert-Jöns et al., *J. of Virology*, Oct. 1991, vol. 65(10): pp. 5184–5189.

Jones et al., *J. of Virology*, Nov. 1991, vol. 65(11): pp. 5860–5872.

Jones et al., *J. of Virology*, Apr. 1991, vol. 65(4): pp. 2024–2036.

Primary Examiner—Mindy Fleisher

Assistant Examiner—Terry A. McKelvey

Attorney, Agent, or Firm—Elizabeth M. Barnhard

[57] **ABSTRACT**

Infection of human fibroblast cells with human cytomegalovirus (HCMV) causes down regulation of cell surface expression of MHC class I. The present invention is directed to a mutant with a 9-kb deletion in the S component of the HCMV genome (including open reading frames IRS1–US9 and US11) which failed to down regulate class I heavy chains. By examining the phenotypes of mutants with smaller deletions with this portion of the HCMV genome, a 7-kb region containing at least 9 open reading frames was shown to contain the genes required for reduction in heavy chain expression. Furthermore, it was determined that two subregions (A and B) of the 7-kb region each contained genes which were sufficient to cause heavy chain down regulation. In subregion B, the US11 gene product is involved. It encodes an endoglycosidase H-sensitive glycoprotein which is intracytoplasmic, similar to the adenovirus type 2 E3-19K glycoprotein which inhibits surface expression of class I heavy chains.

4 Claims, 17 Drawing Sheets

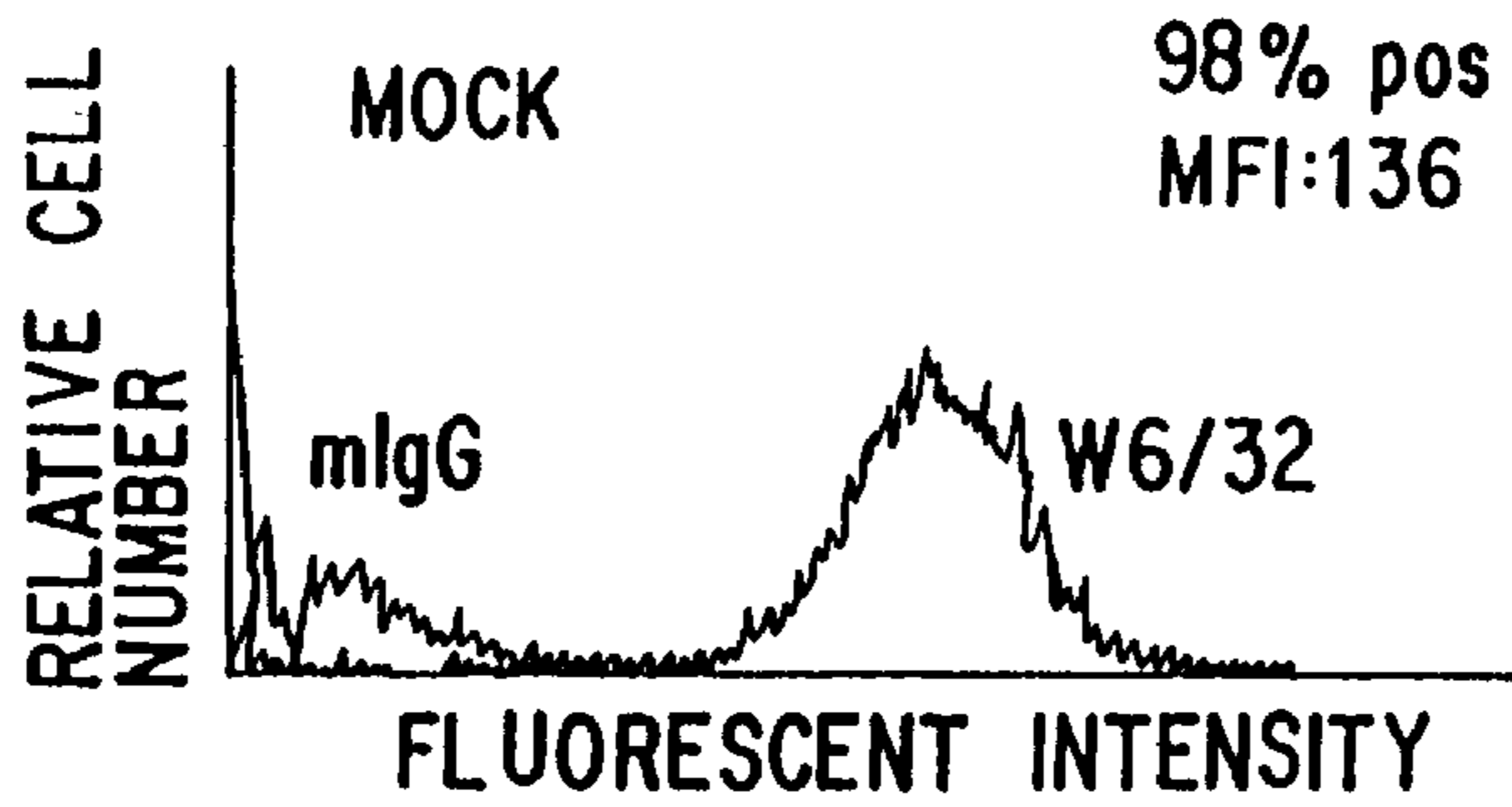


FIG. 1A

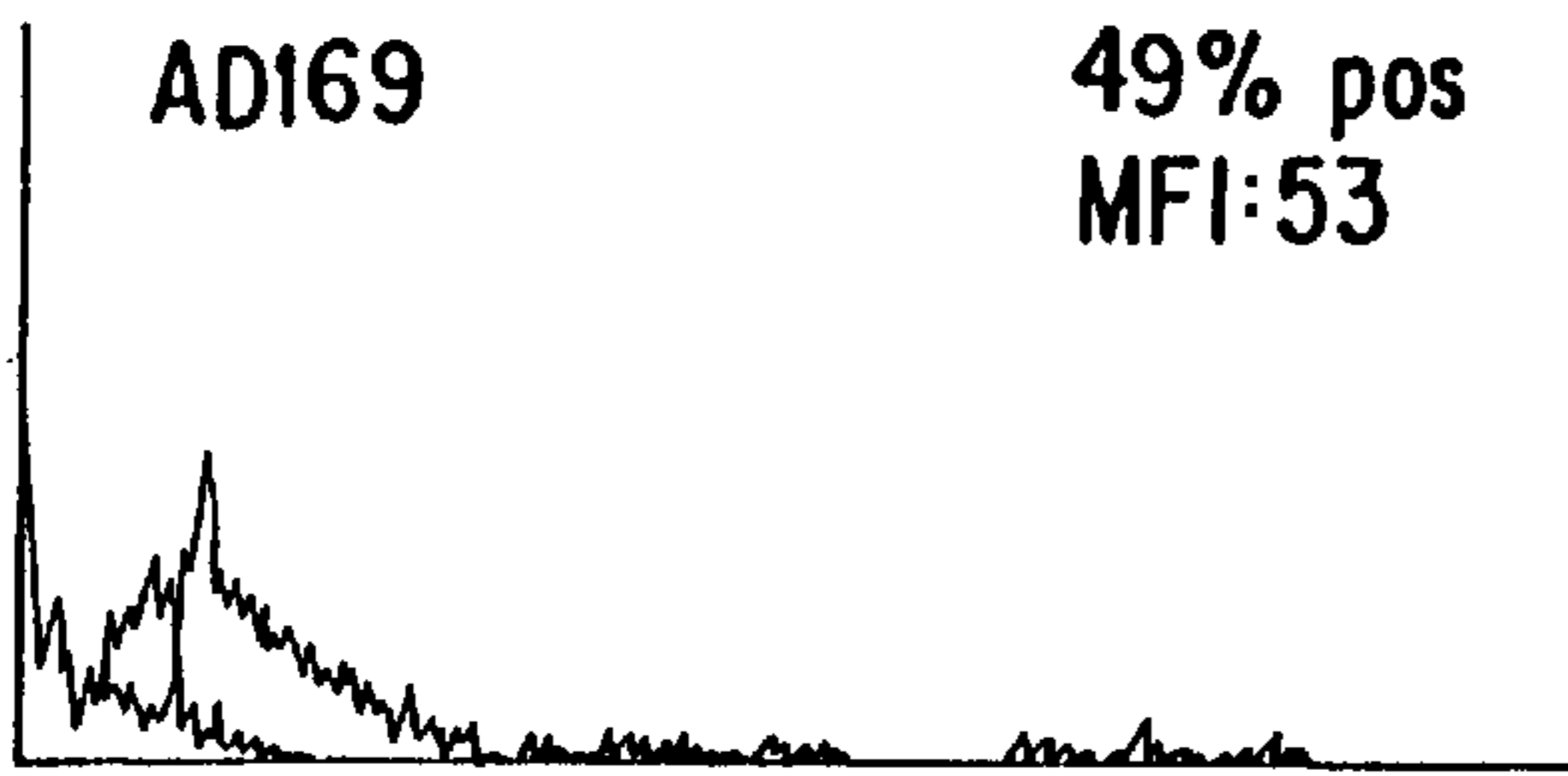


FIG. 1B

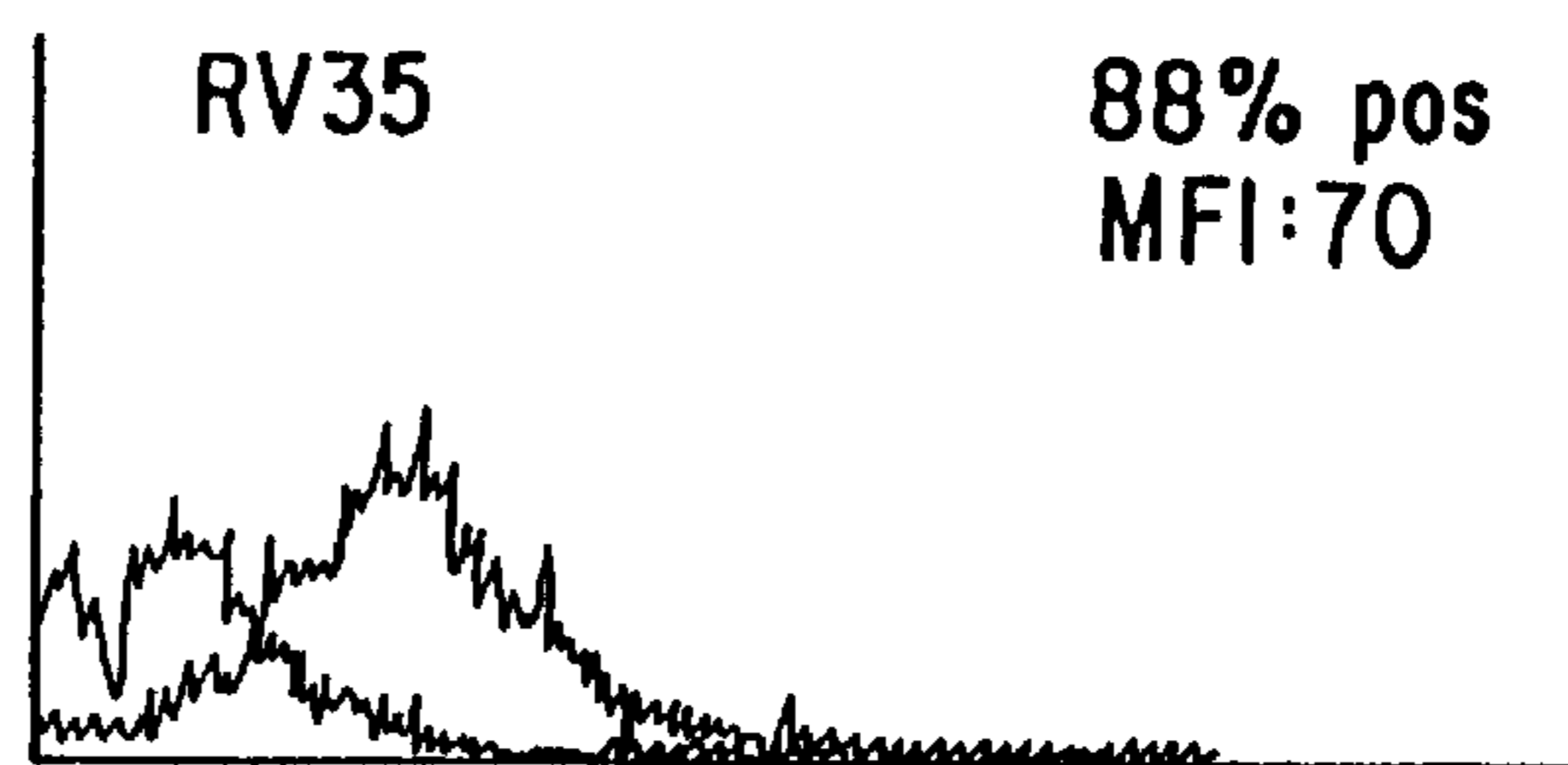


FIG. 1C

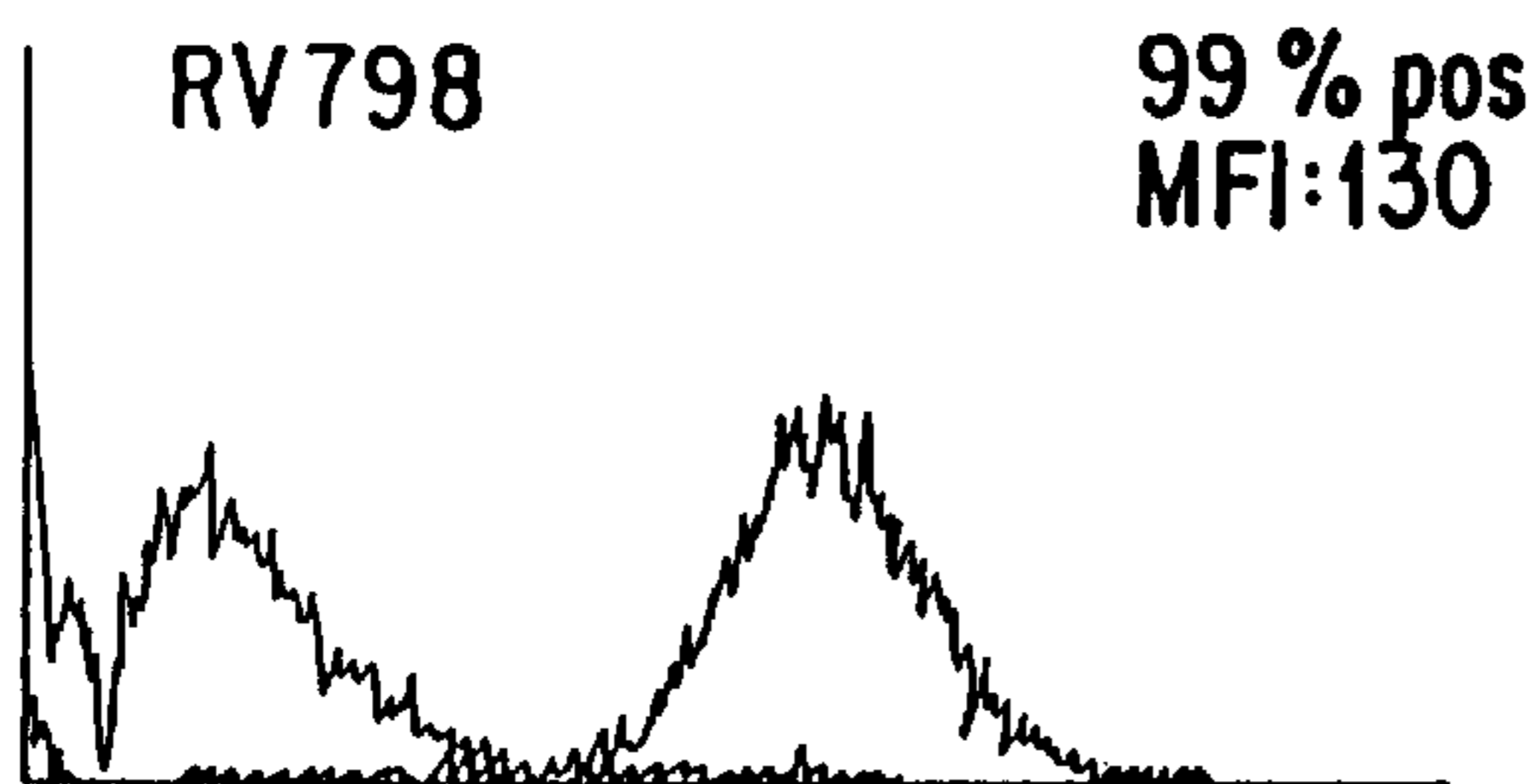


FIG. 1D

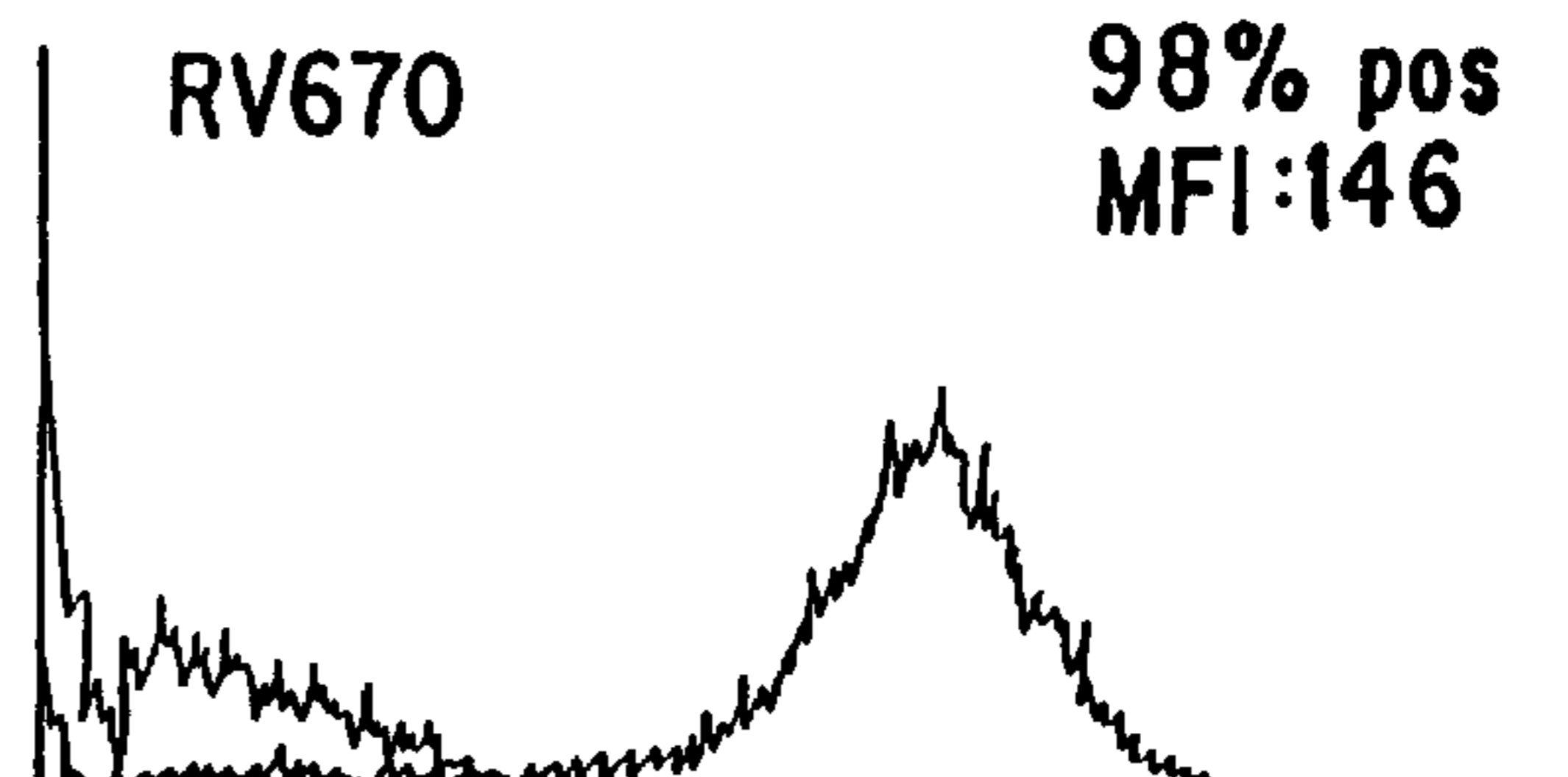


FIG. 1E

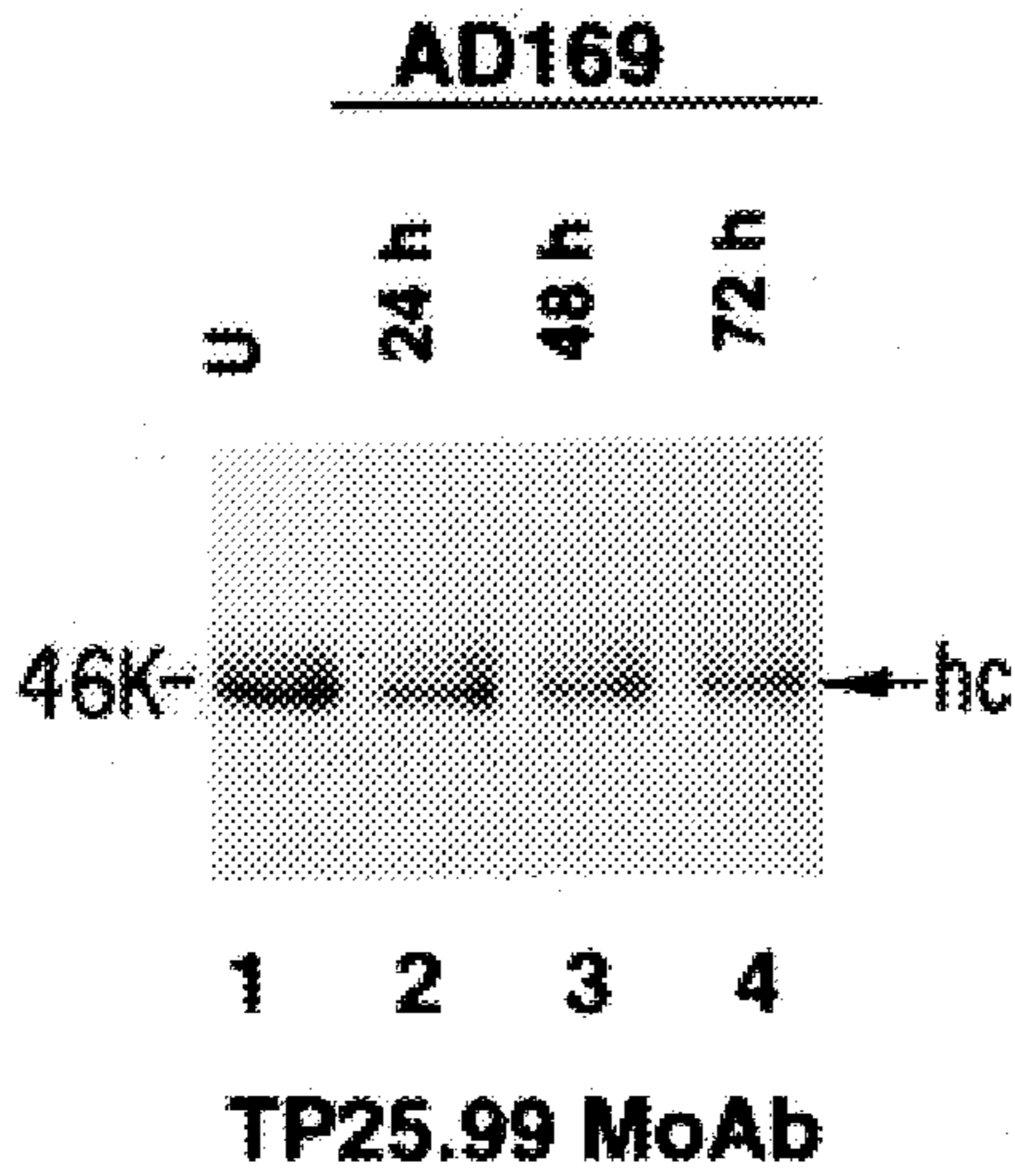


FIG. 2A

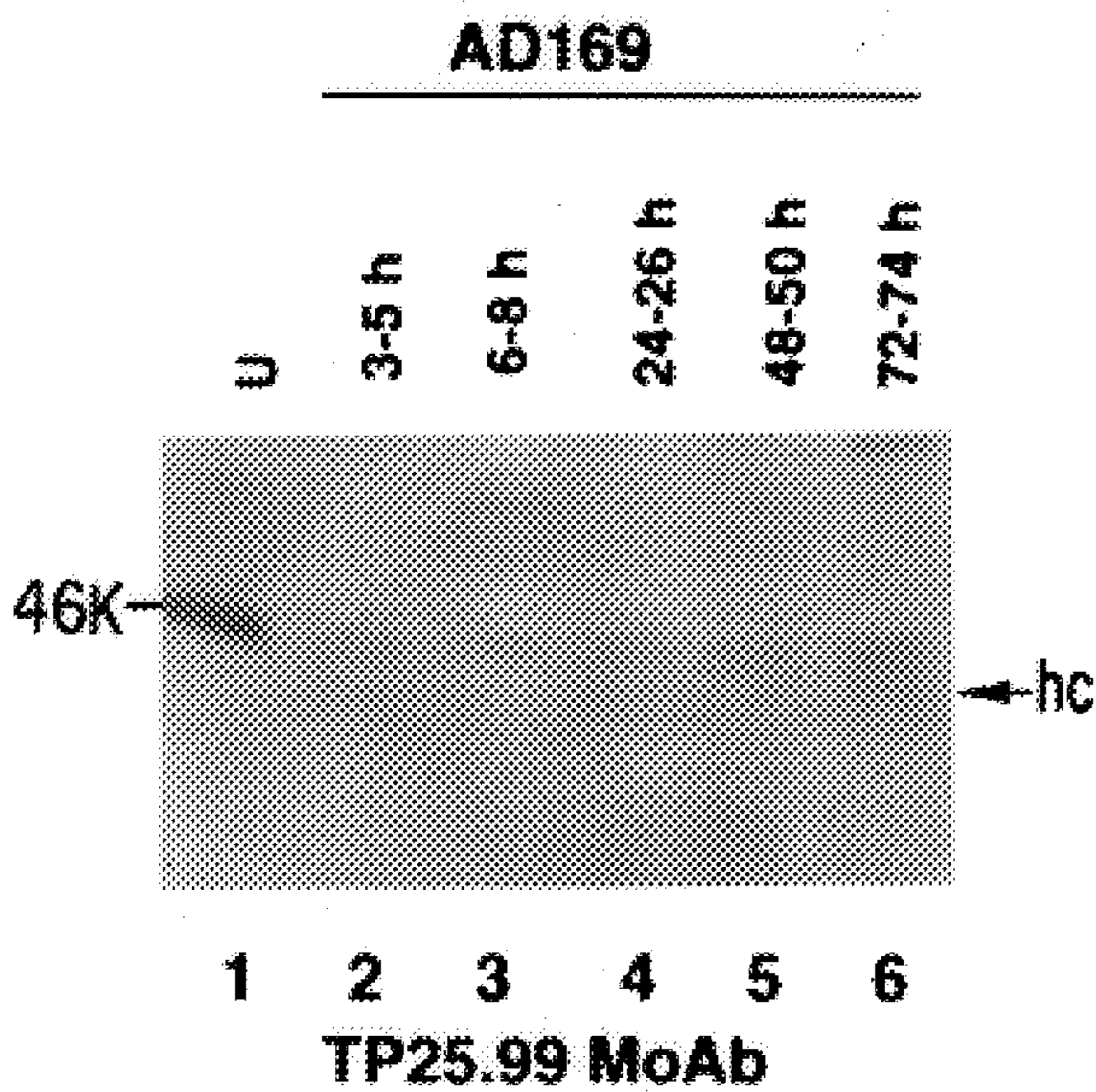


FIG. 2C

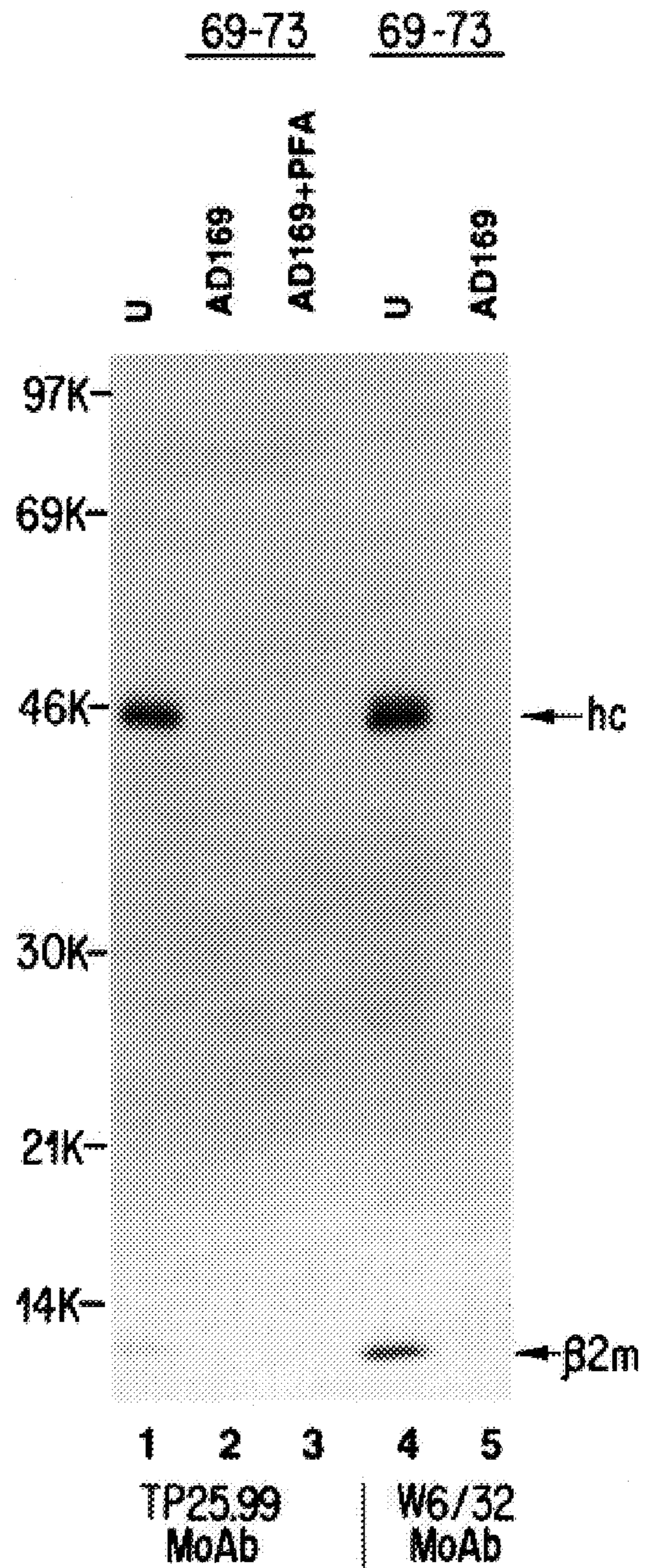


FIG. 2B

HETEROGENEOUS
JOINT REGION
A SEQUENCE

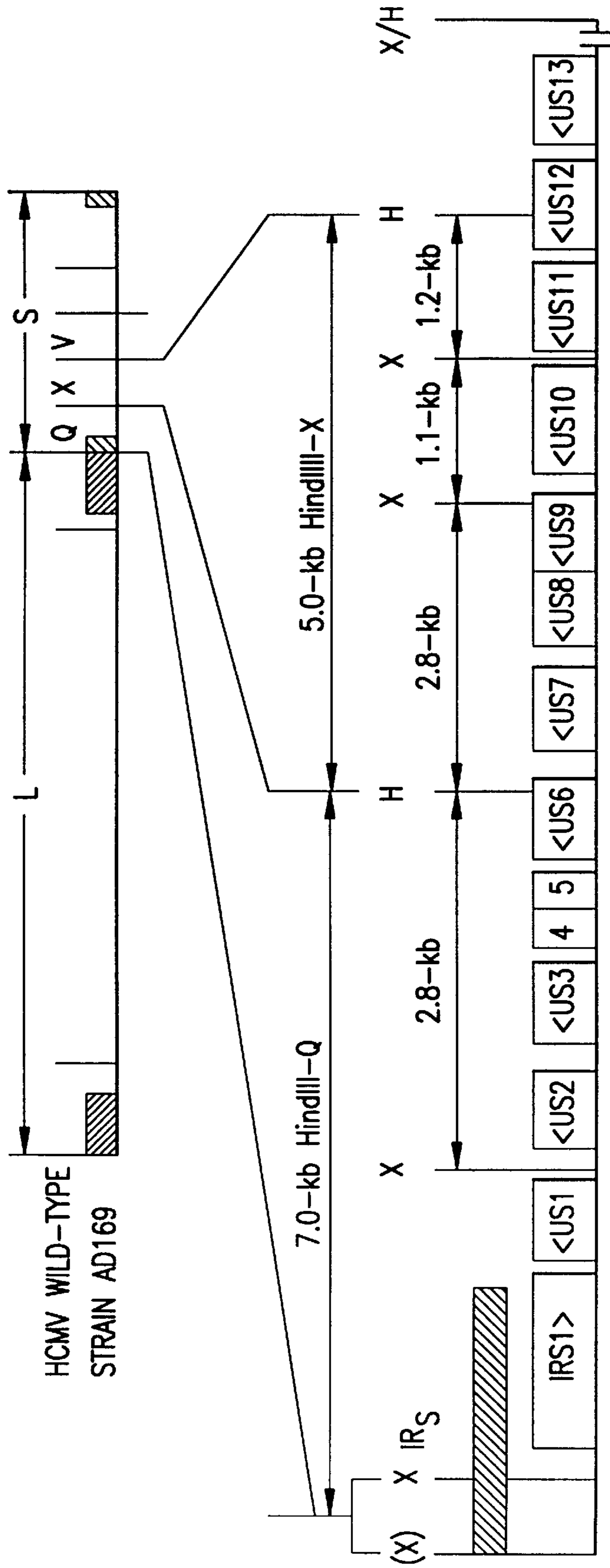
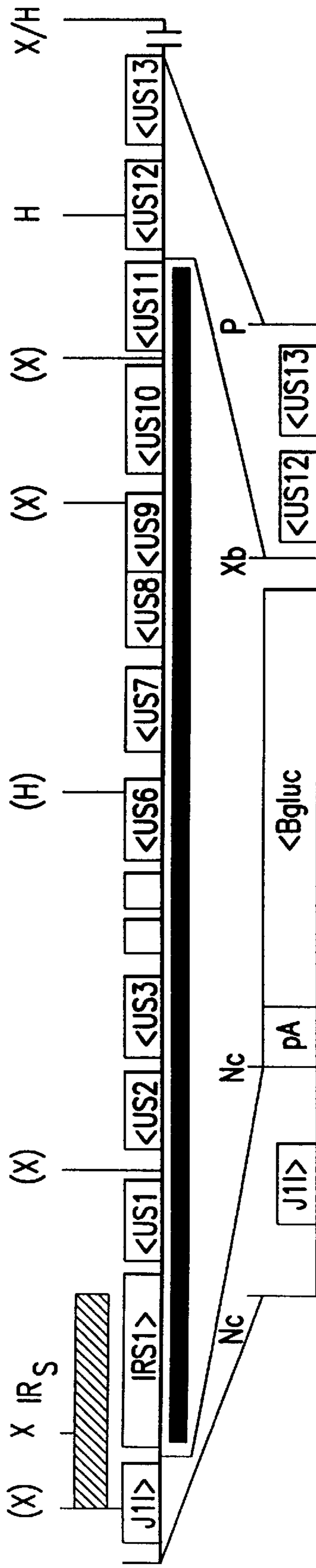
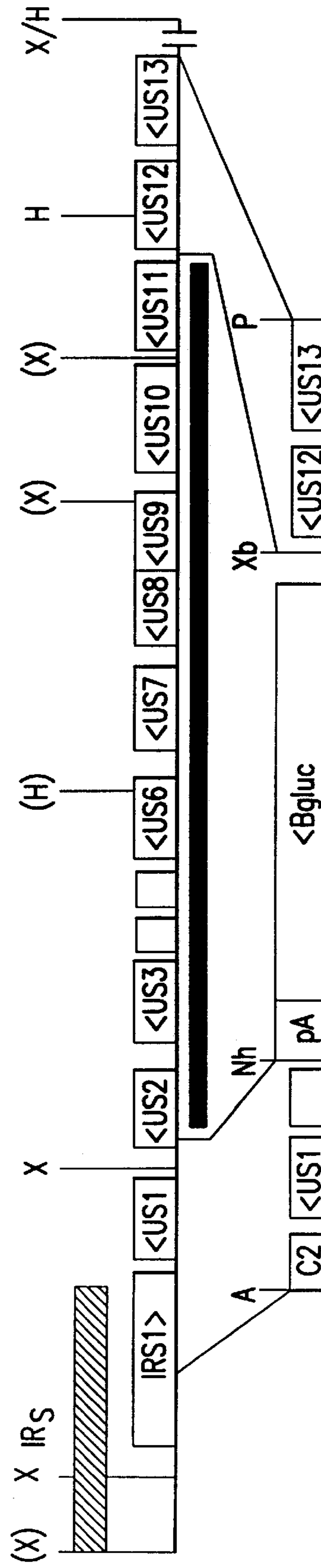


FIG. 3A



RV7186 (IRS1-US11 DELETED)

FIG. 3B



RV798 (US2-US11 DELETED)

FIG. 3C

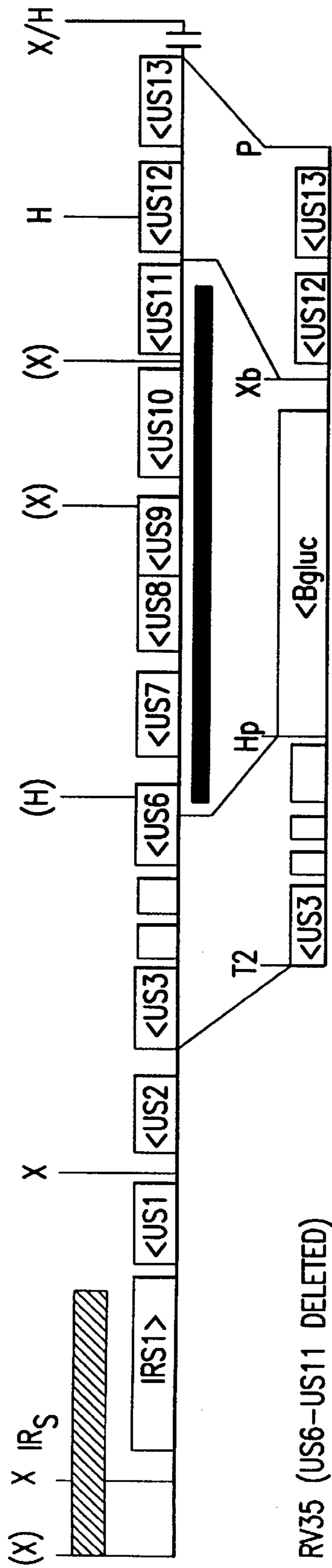


FIG. 3D

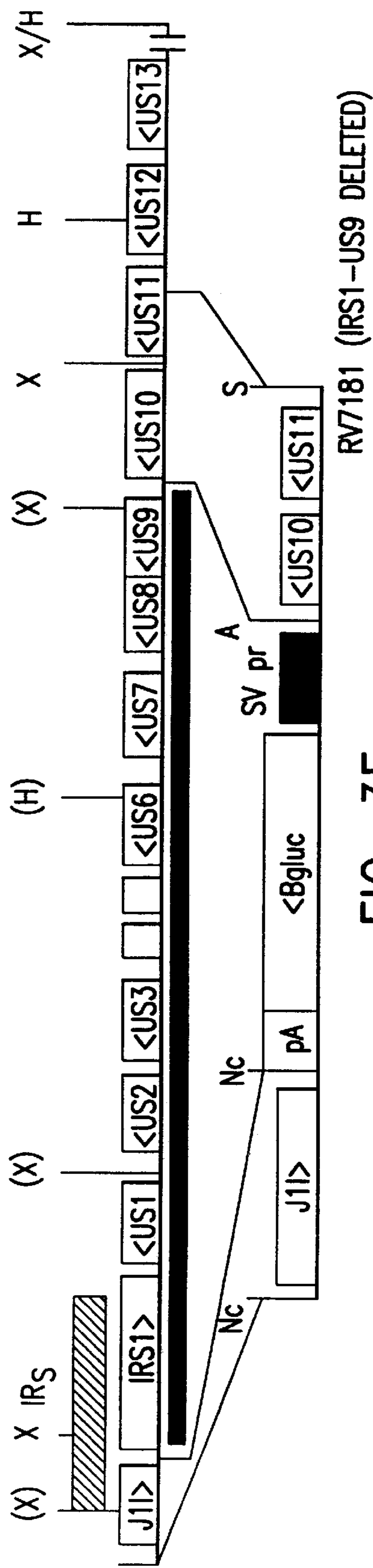


FIG. 3E

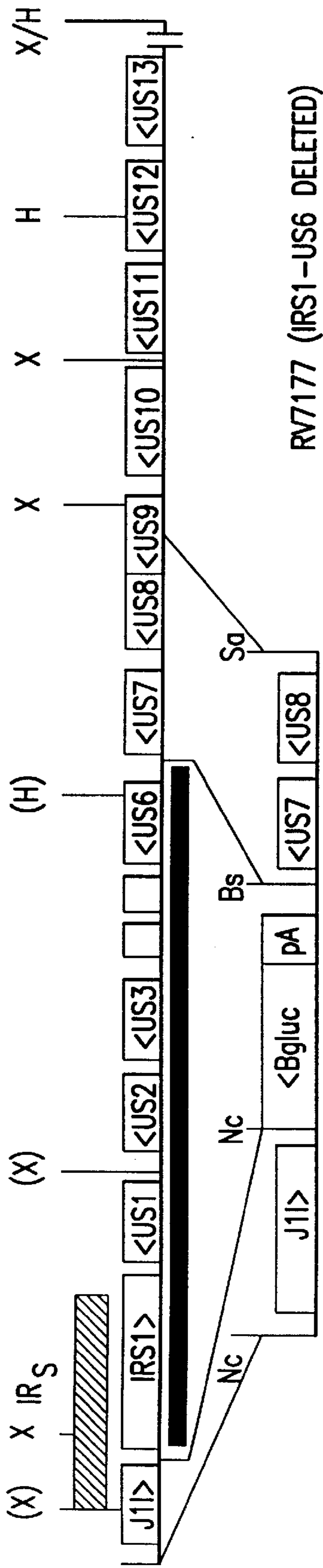


FIG. 3F

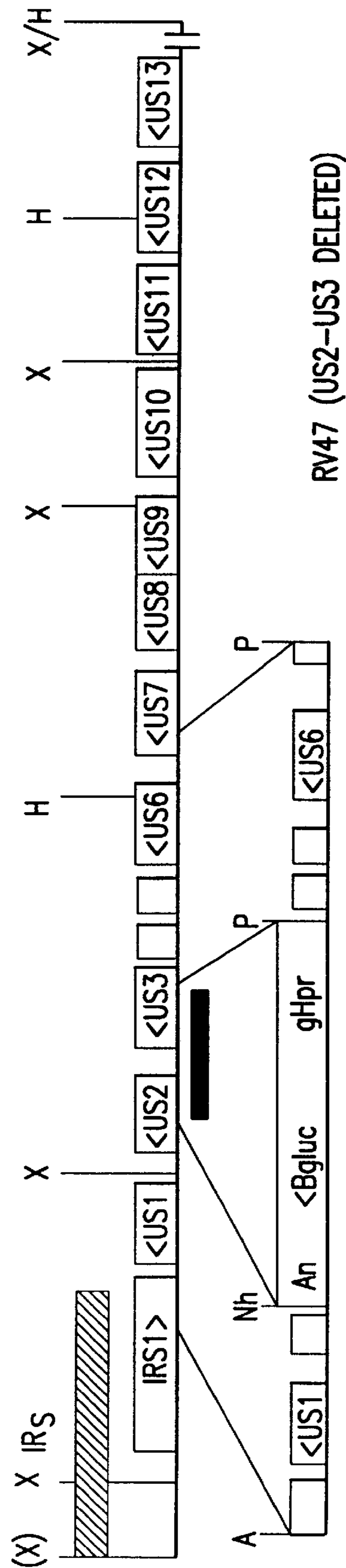
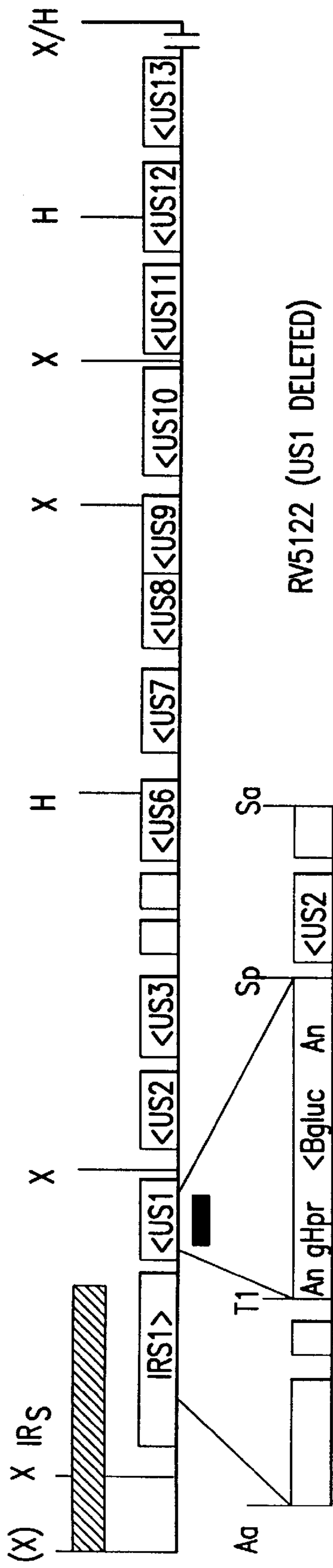
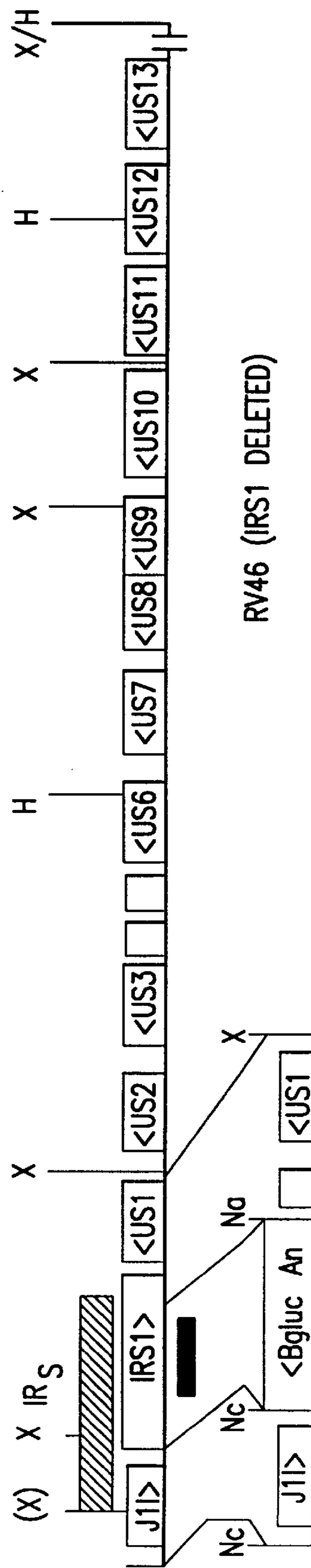


FIG. 3G



RV5122 (US1 DELETED)

FIG. 3H



RV46 (IRS1 DELETED)

FIG. 3I

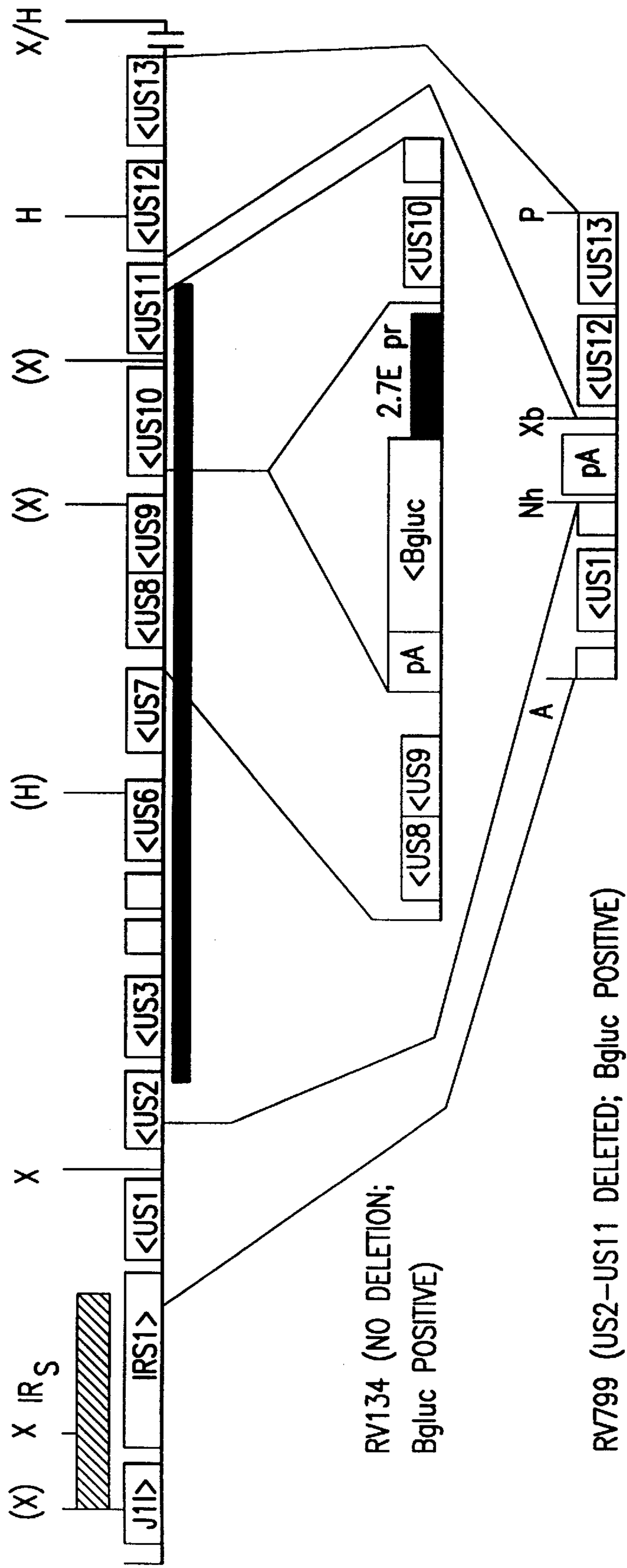
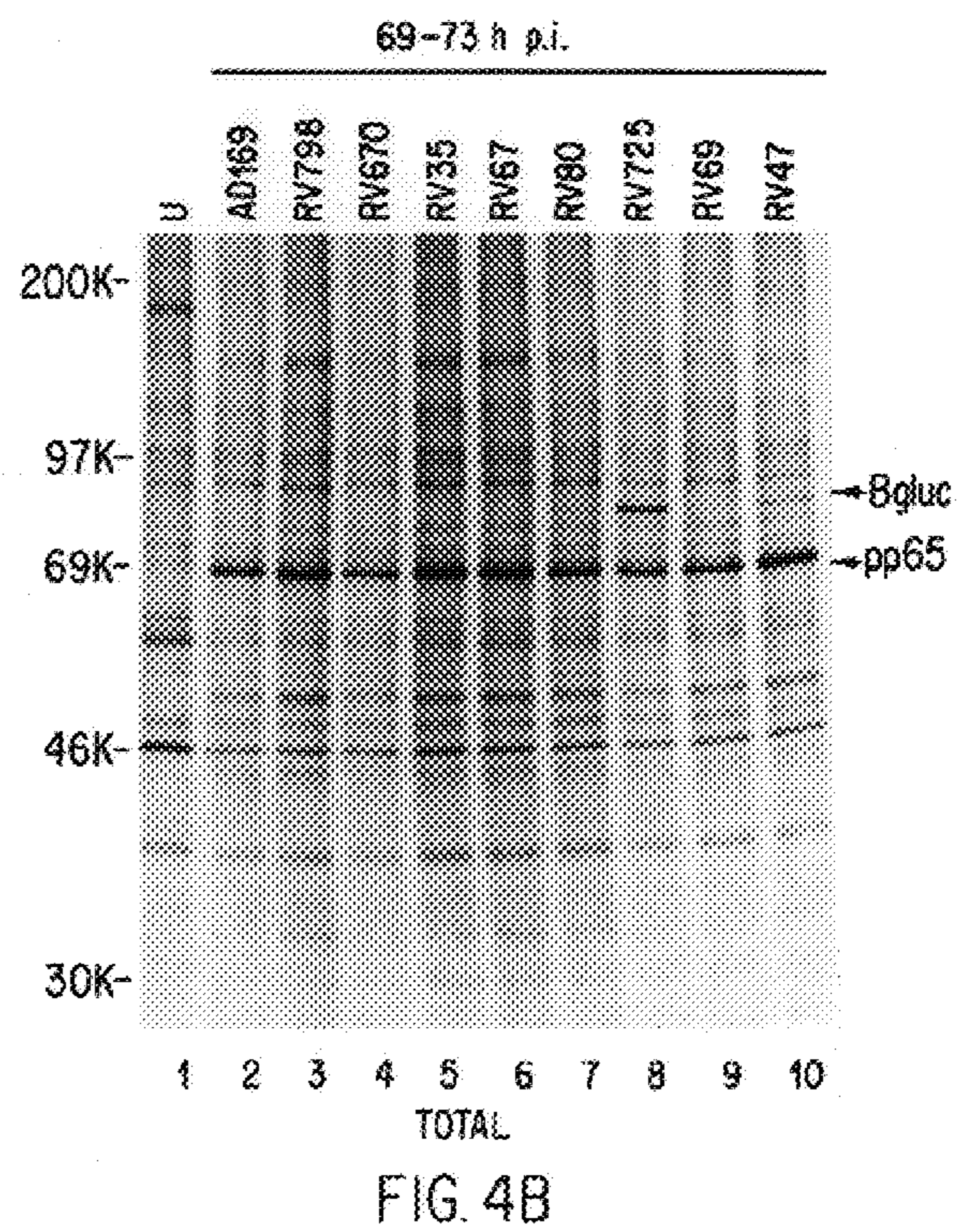
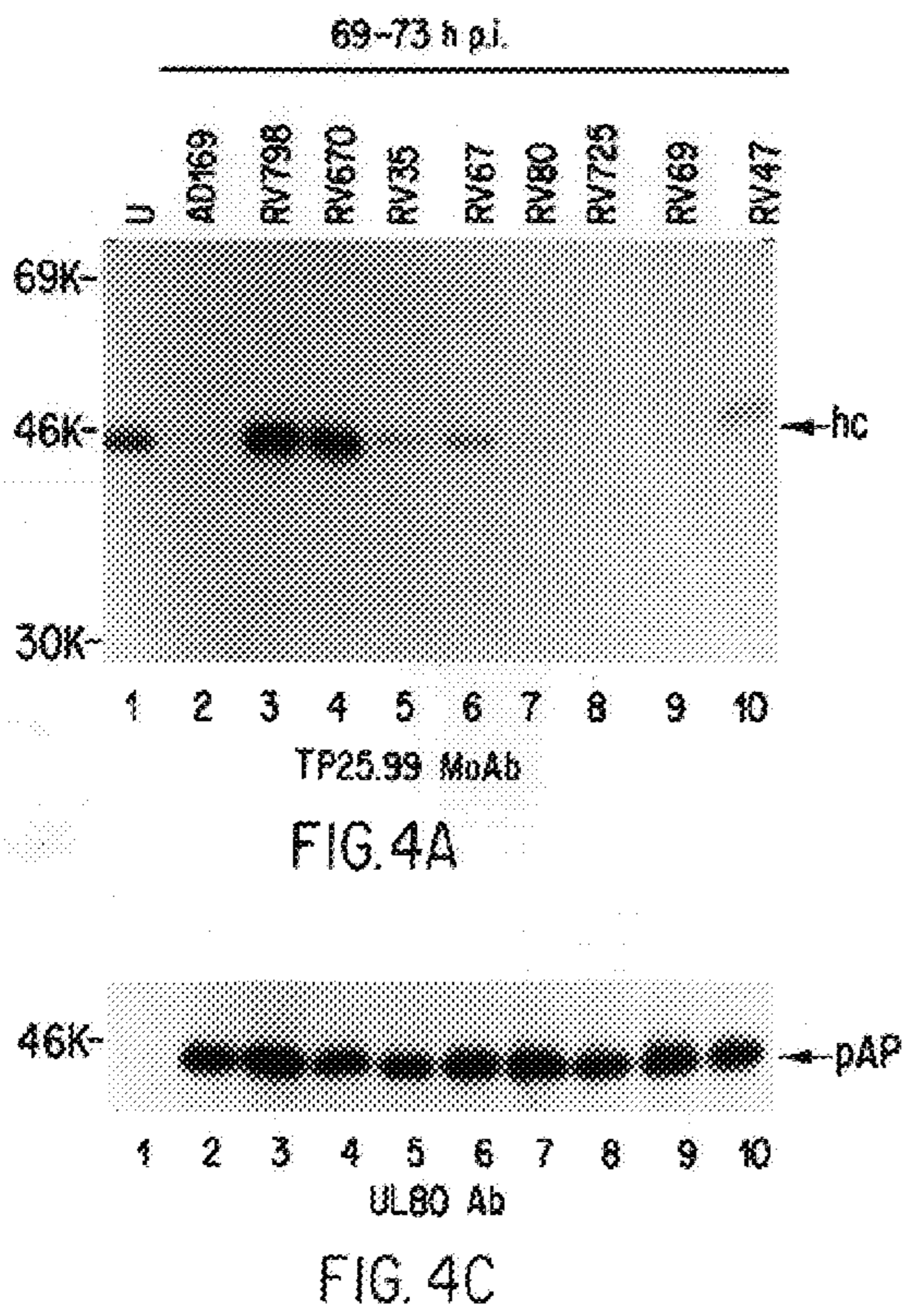


FIG. 3J



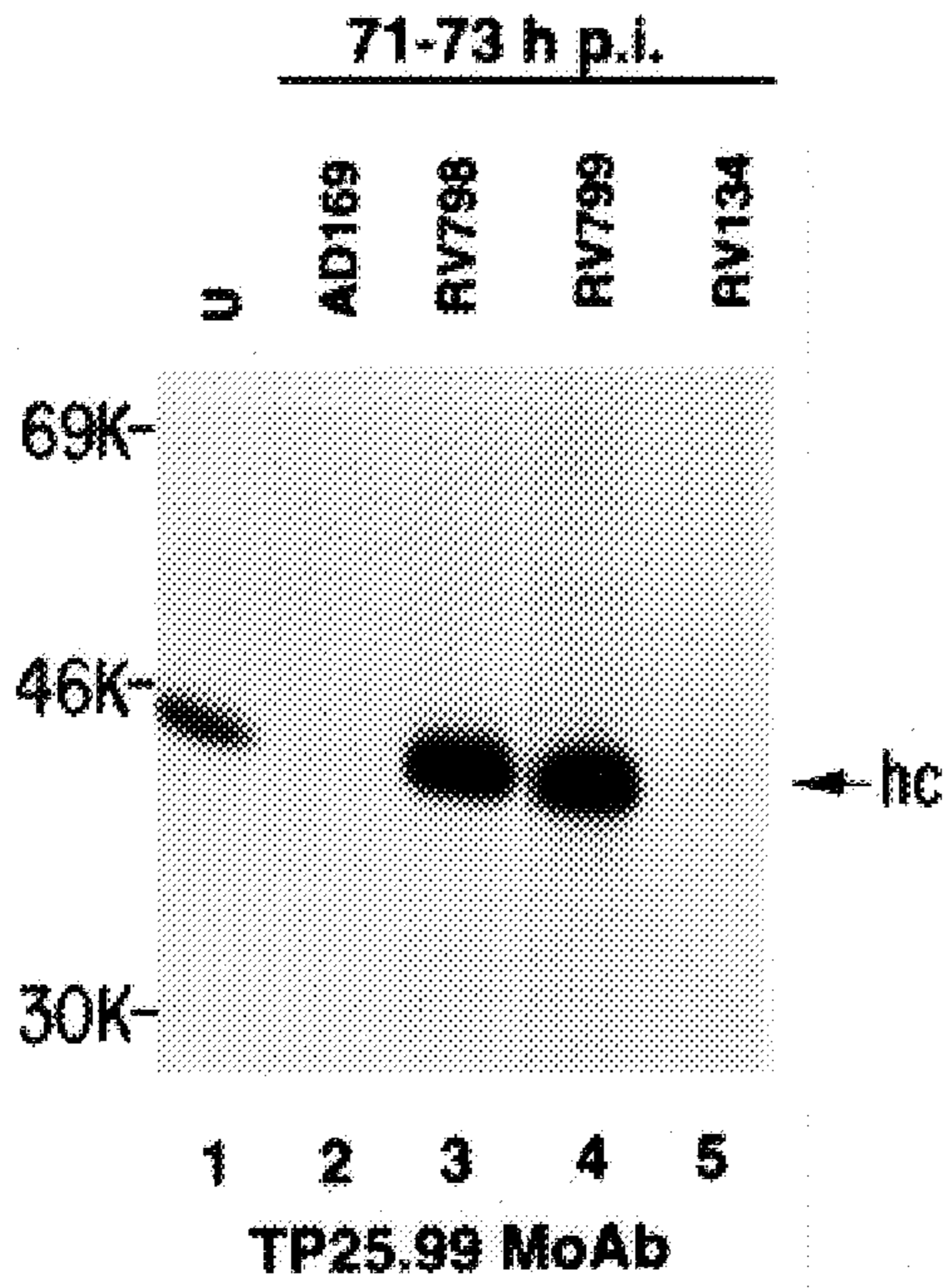


FIG. 5A



FIG. 5C

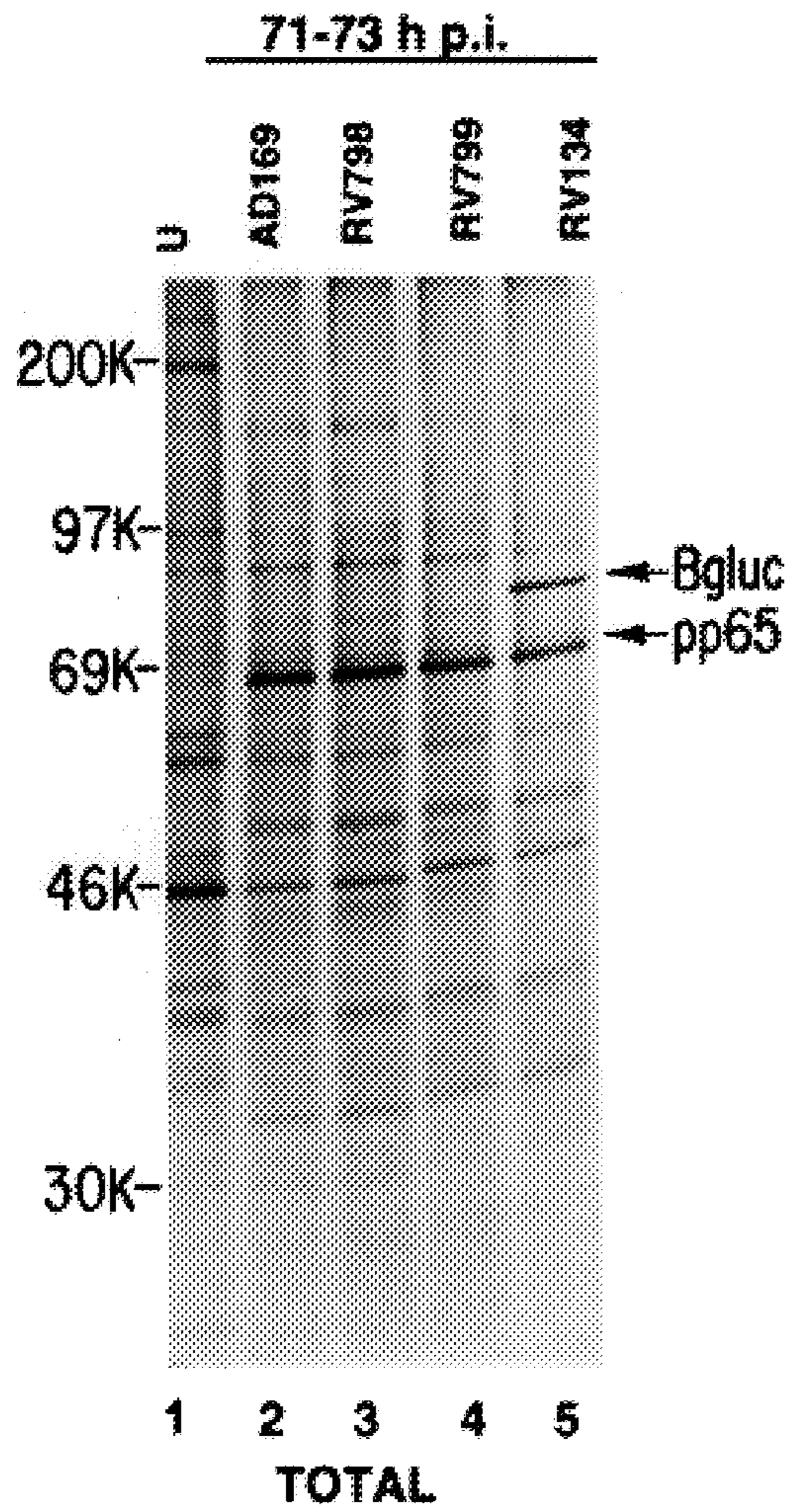


FIG. 5B

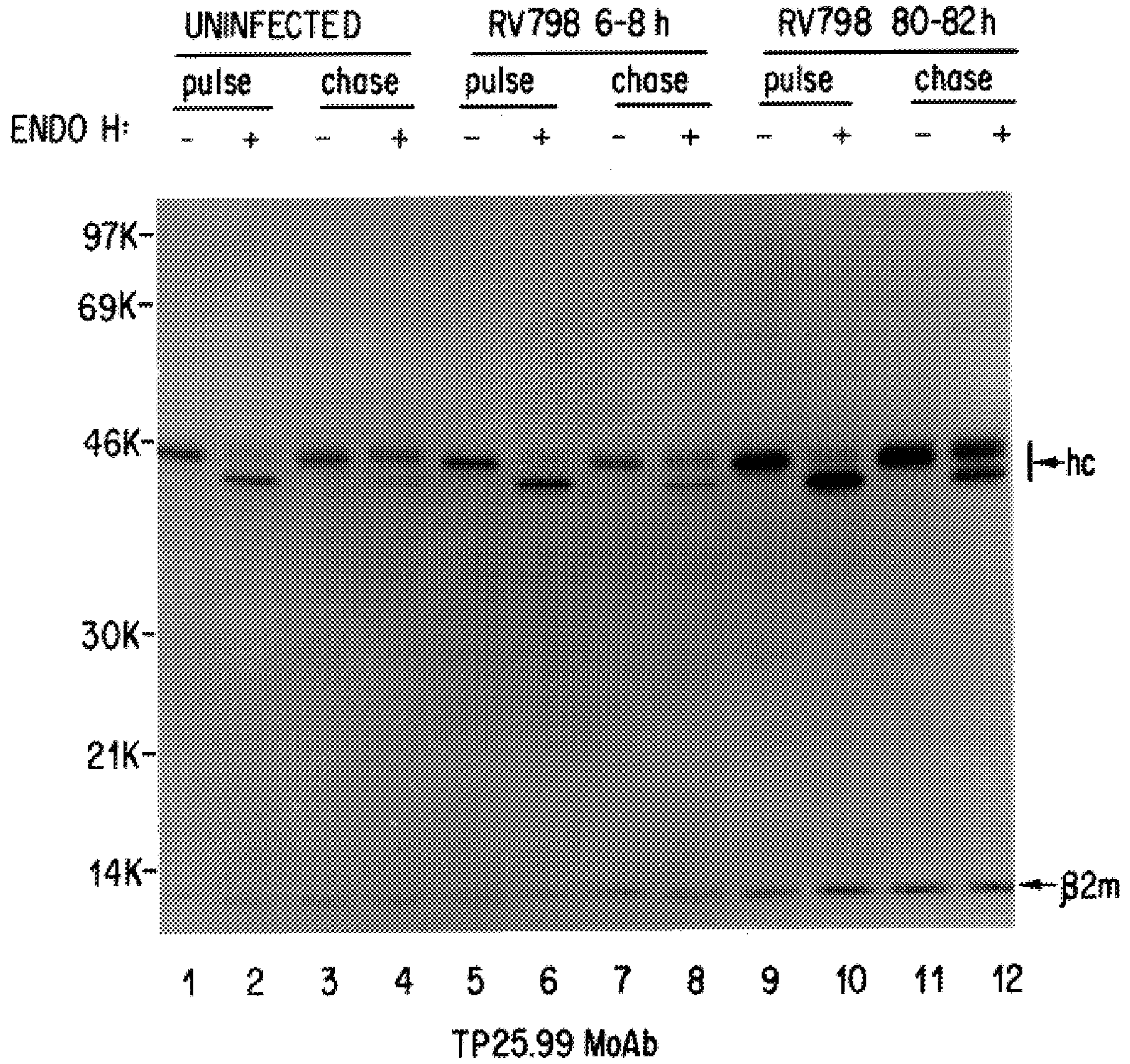


FIG. 6

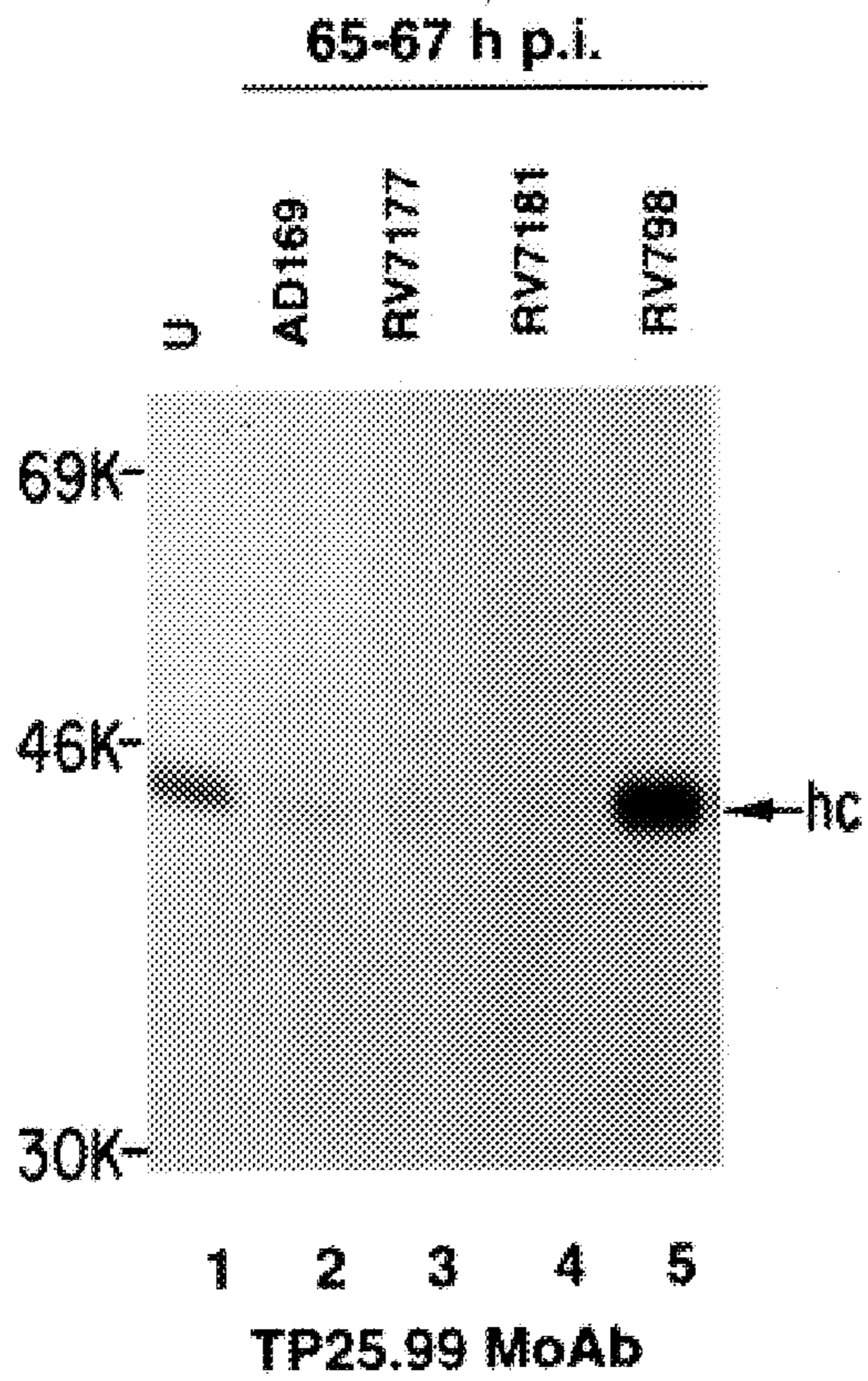


FIG. 7A

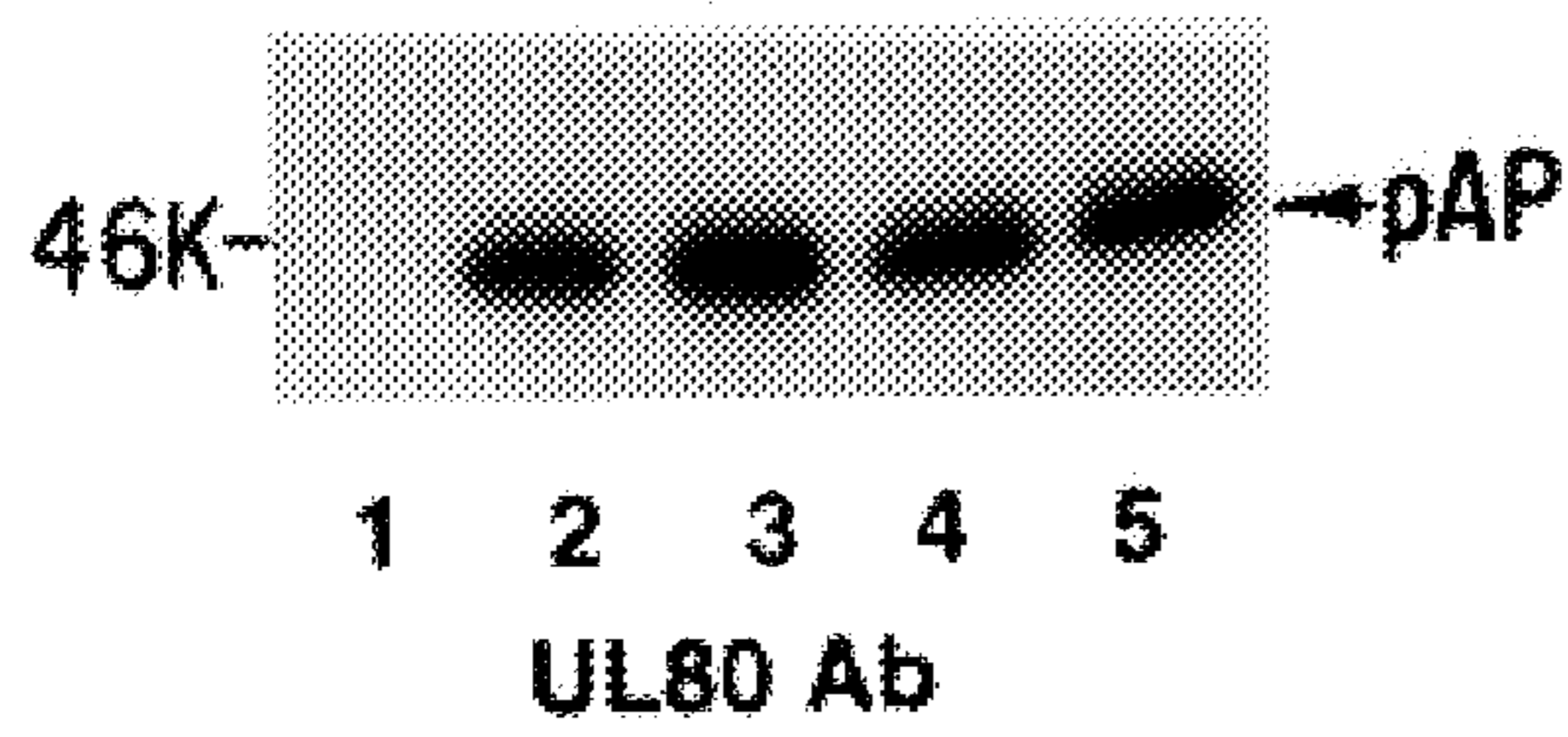


FIG. 7C

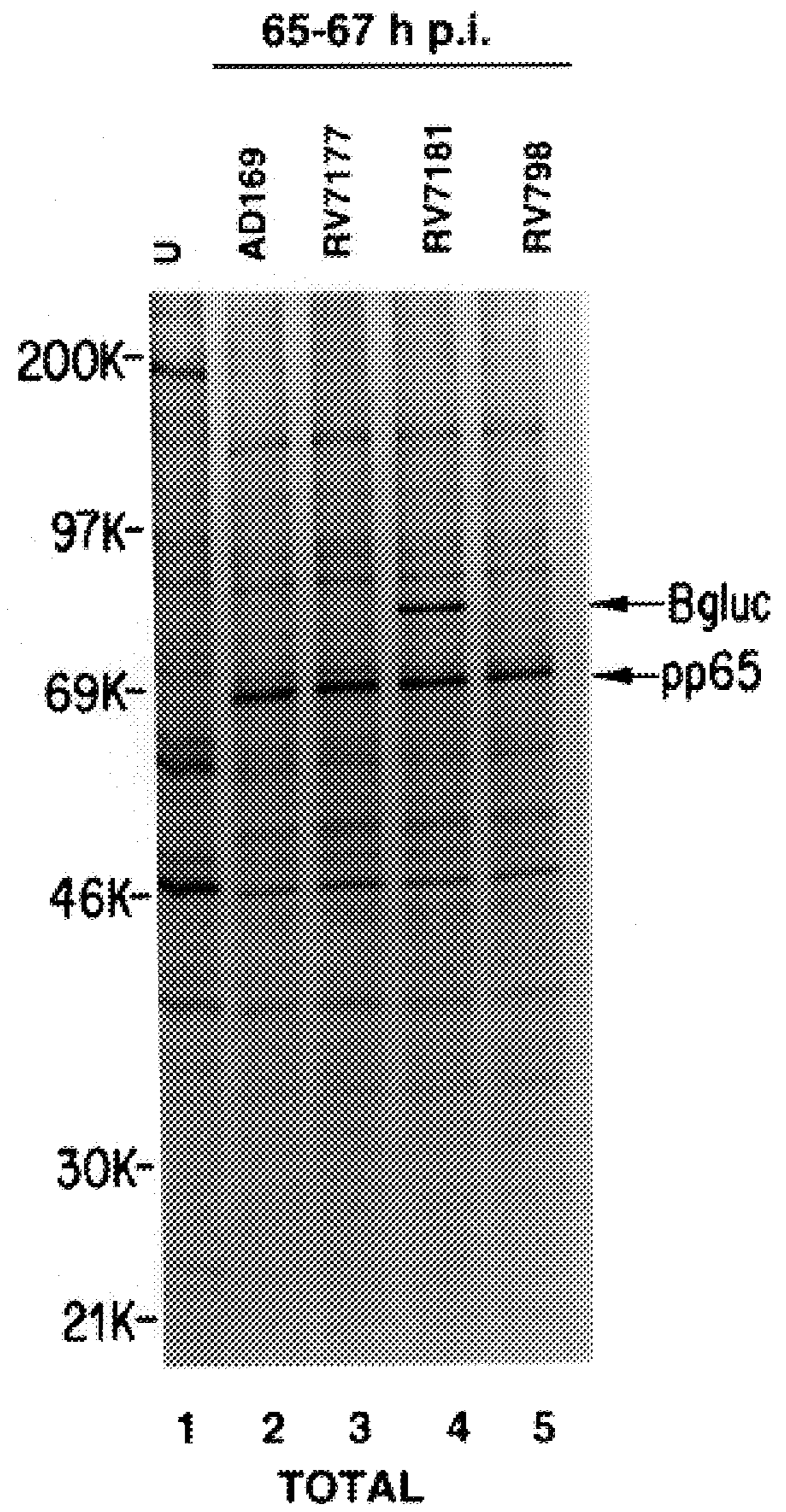


FIG. 7B

PERMEABILIZED

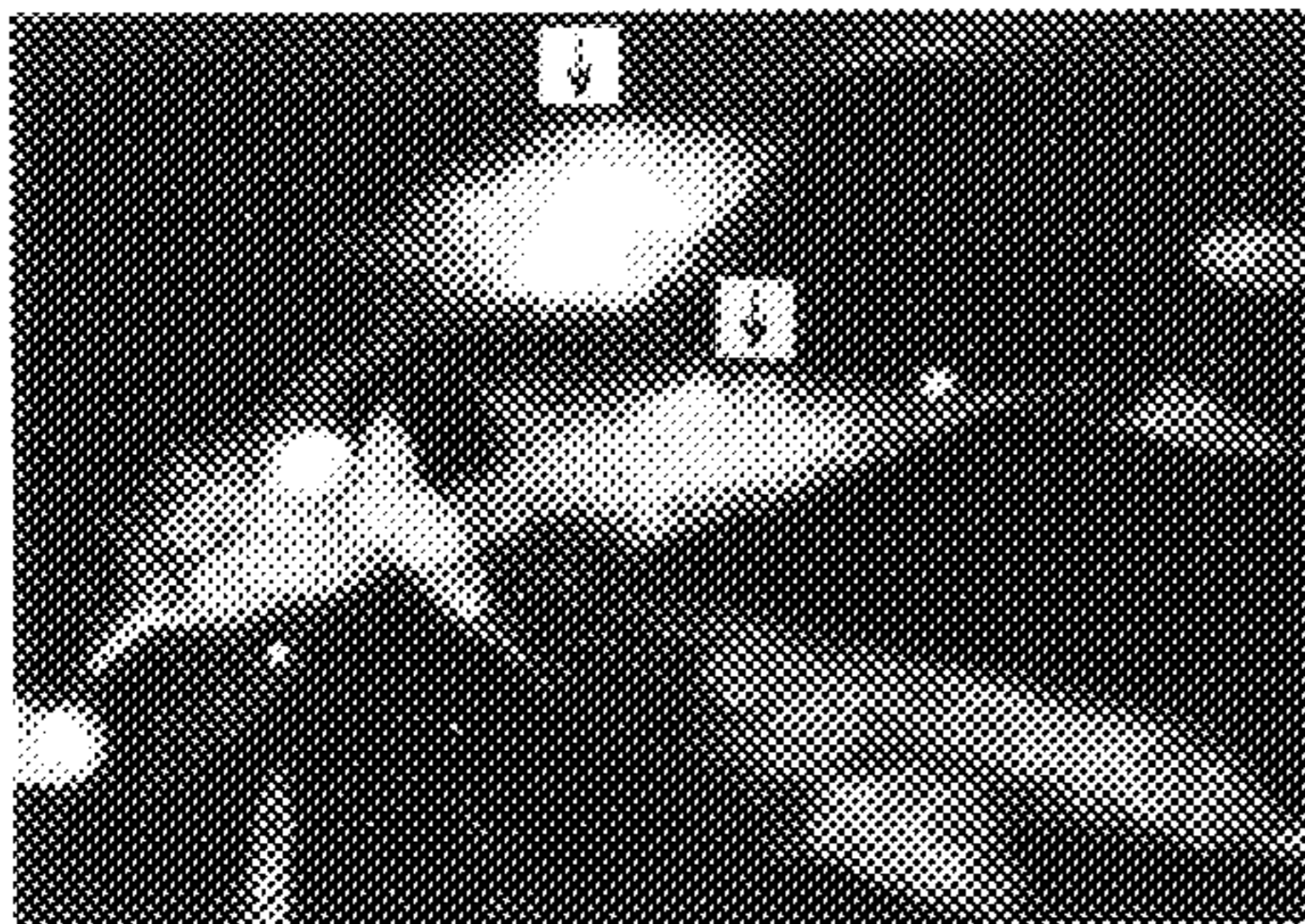


FIG.8A

AD169 (WT) 8h p.i.

NONPERMEABILIZED

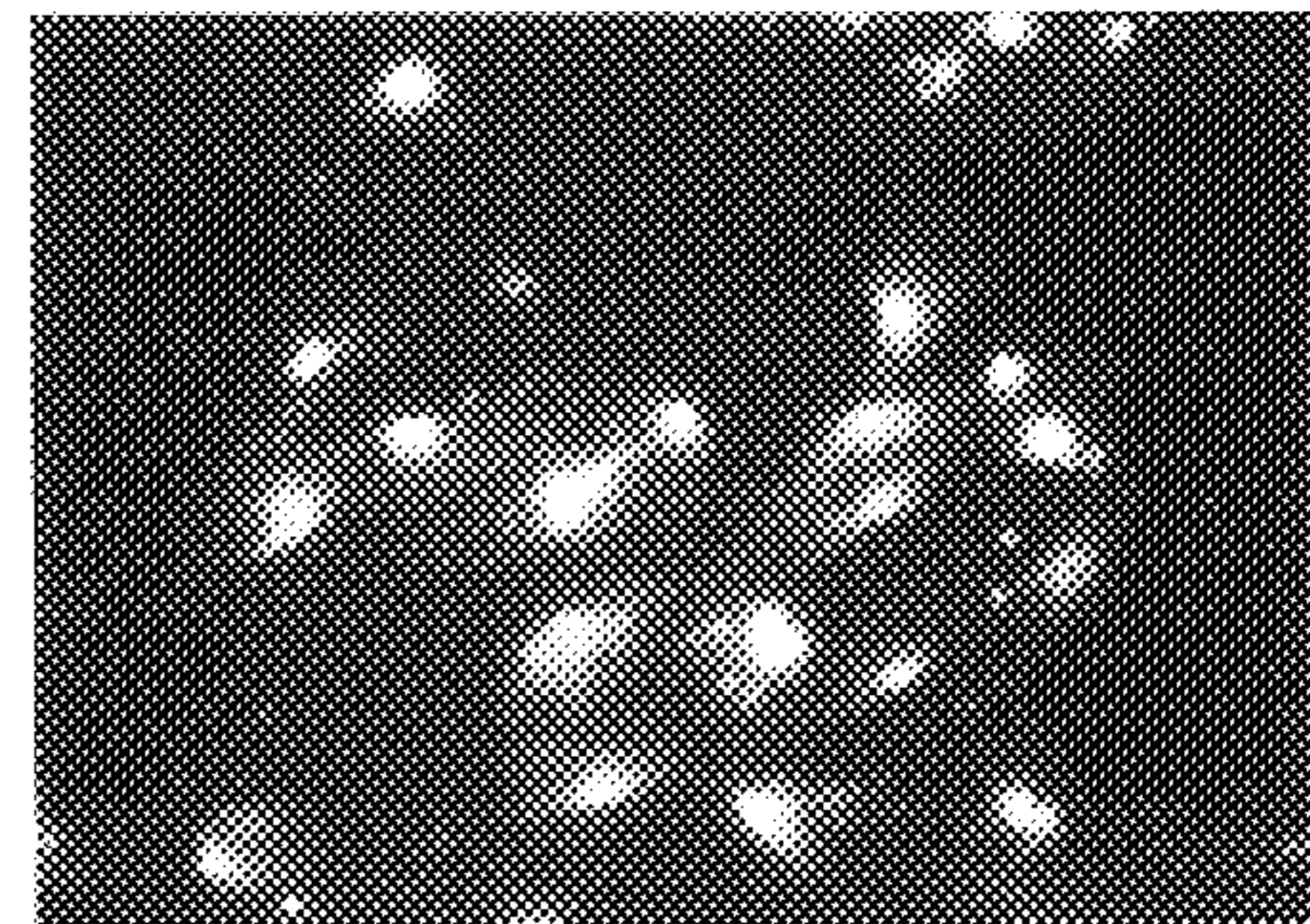


FIG.8B

AD169 (WT) 8h p.i.

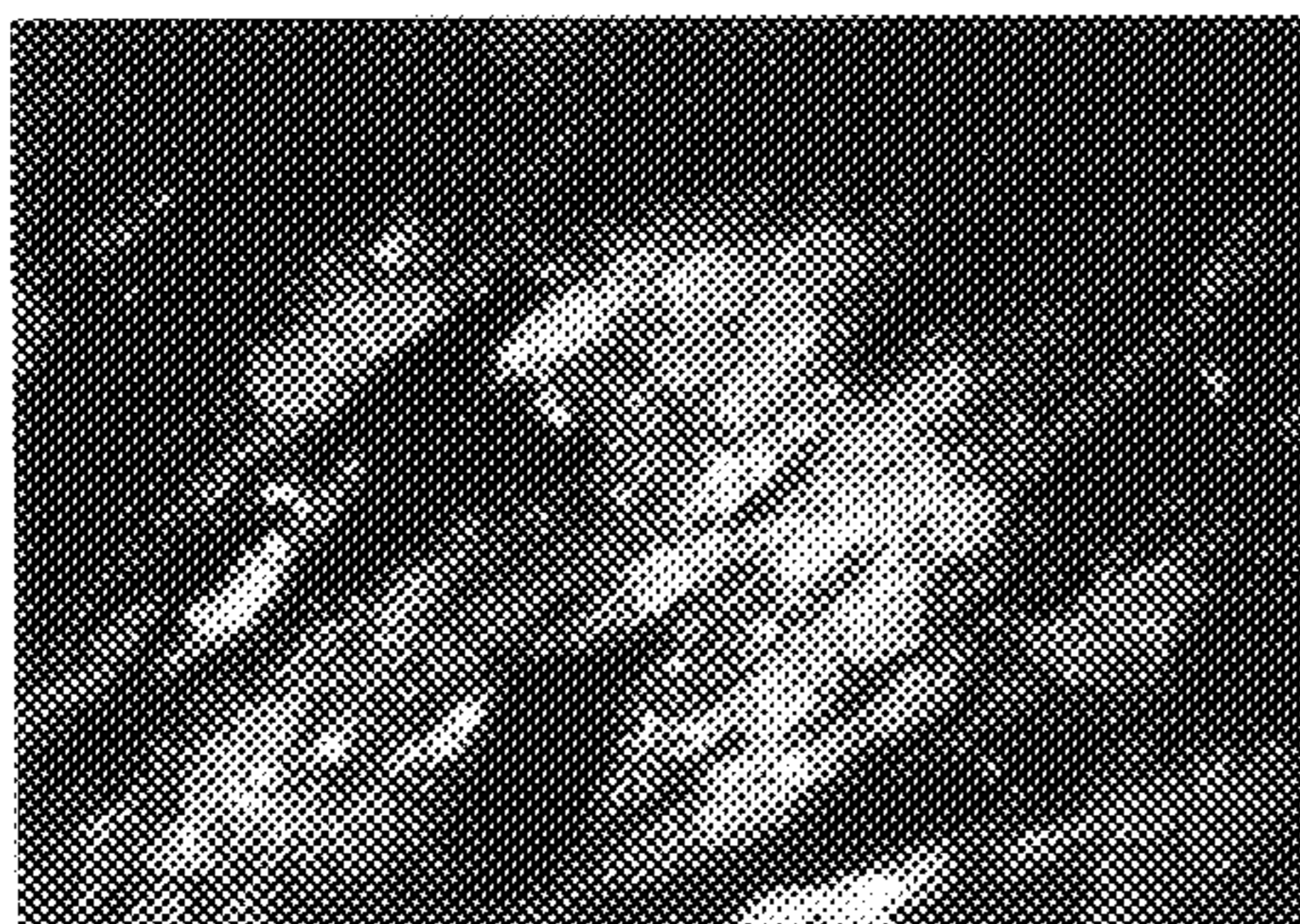


FIG.8C

RV699 8h p.i.

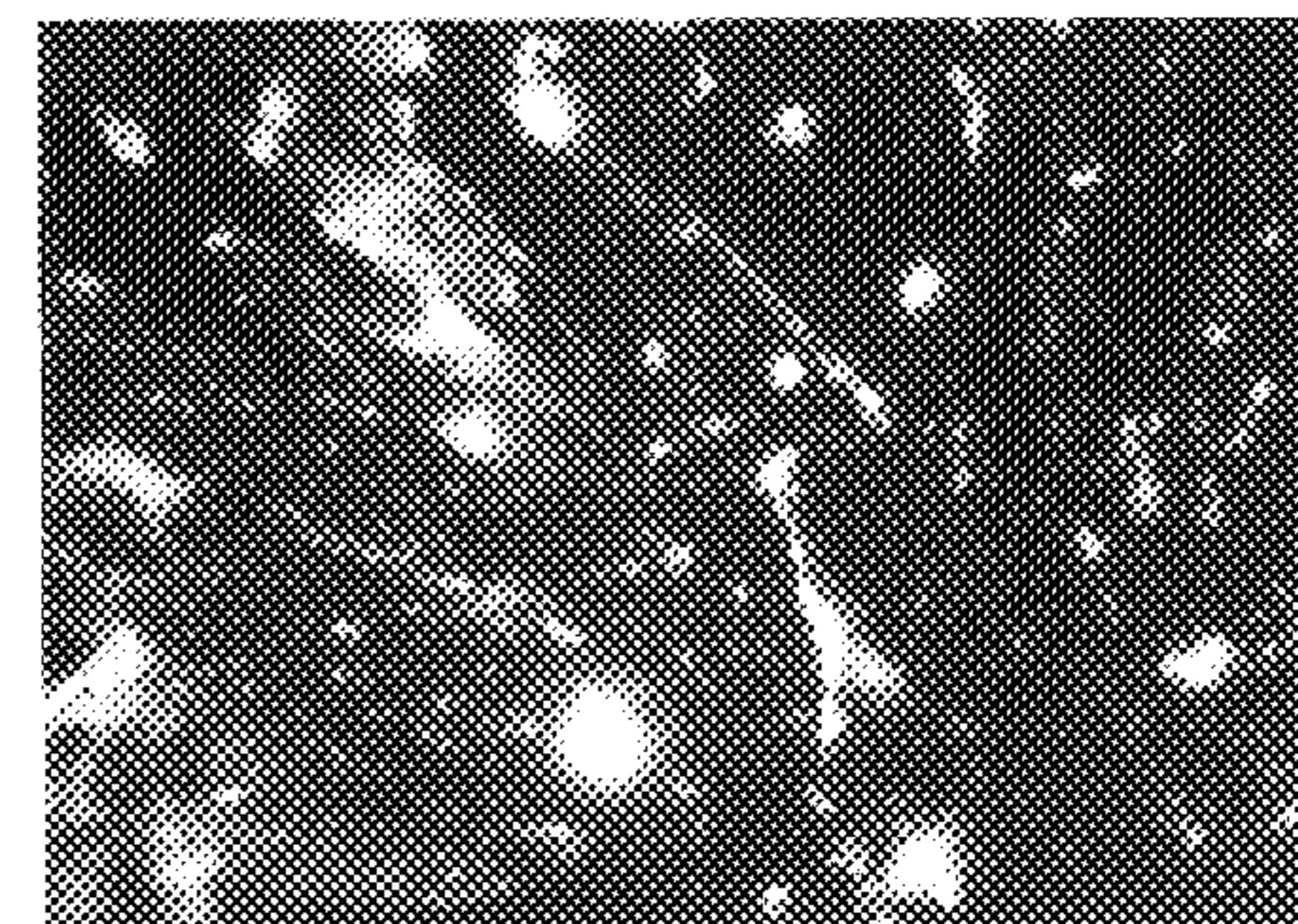


FIG.8D

RV699 8h p.i.

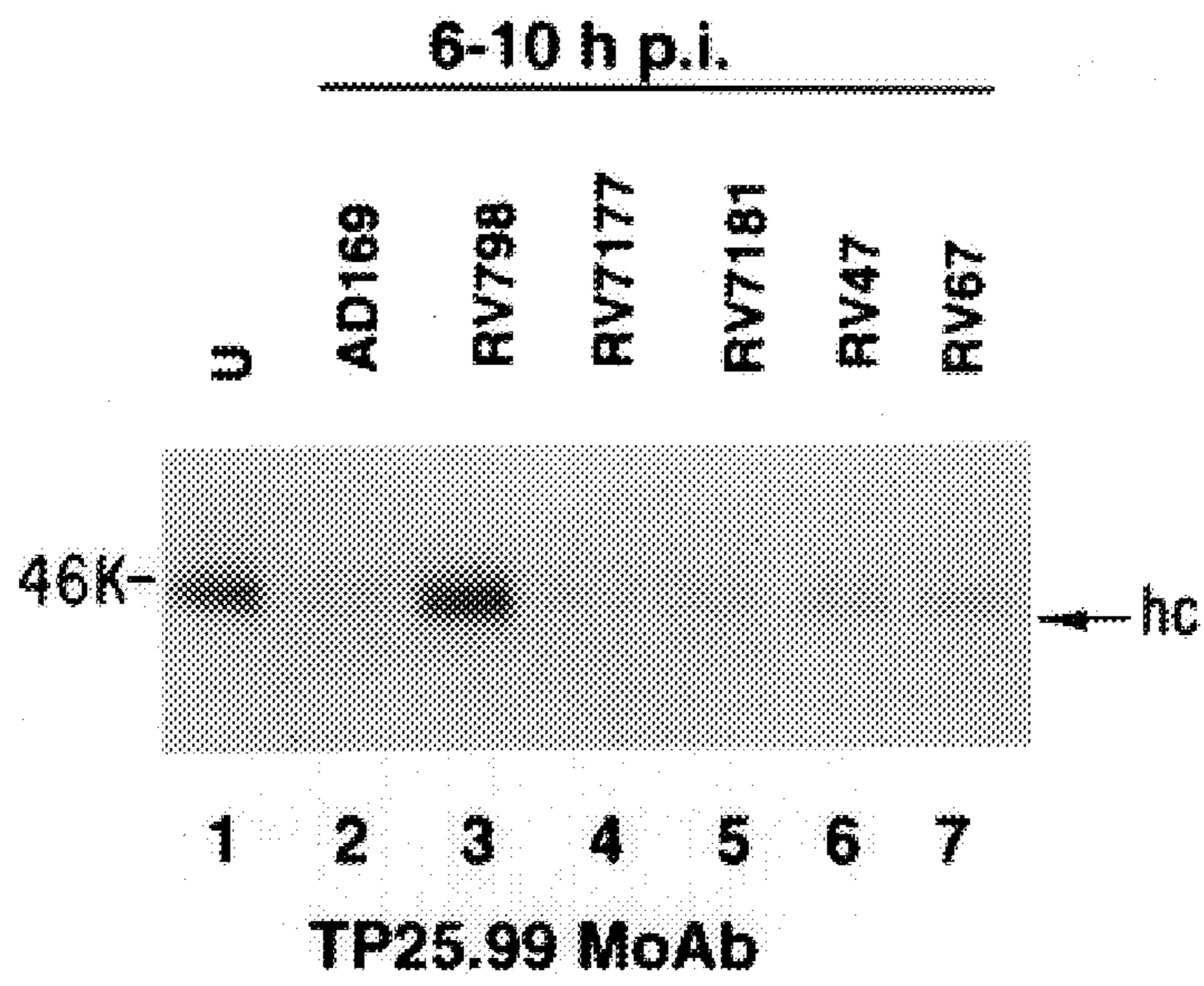


FIG. 9A

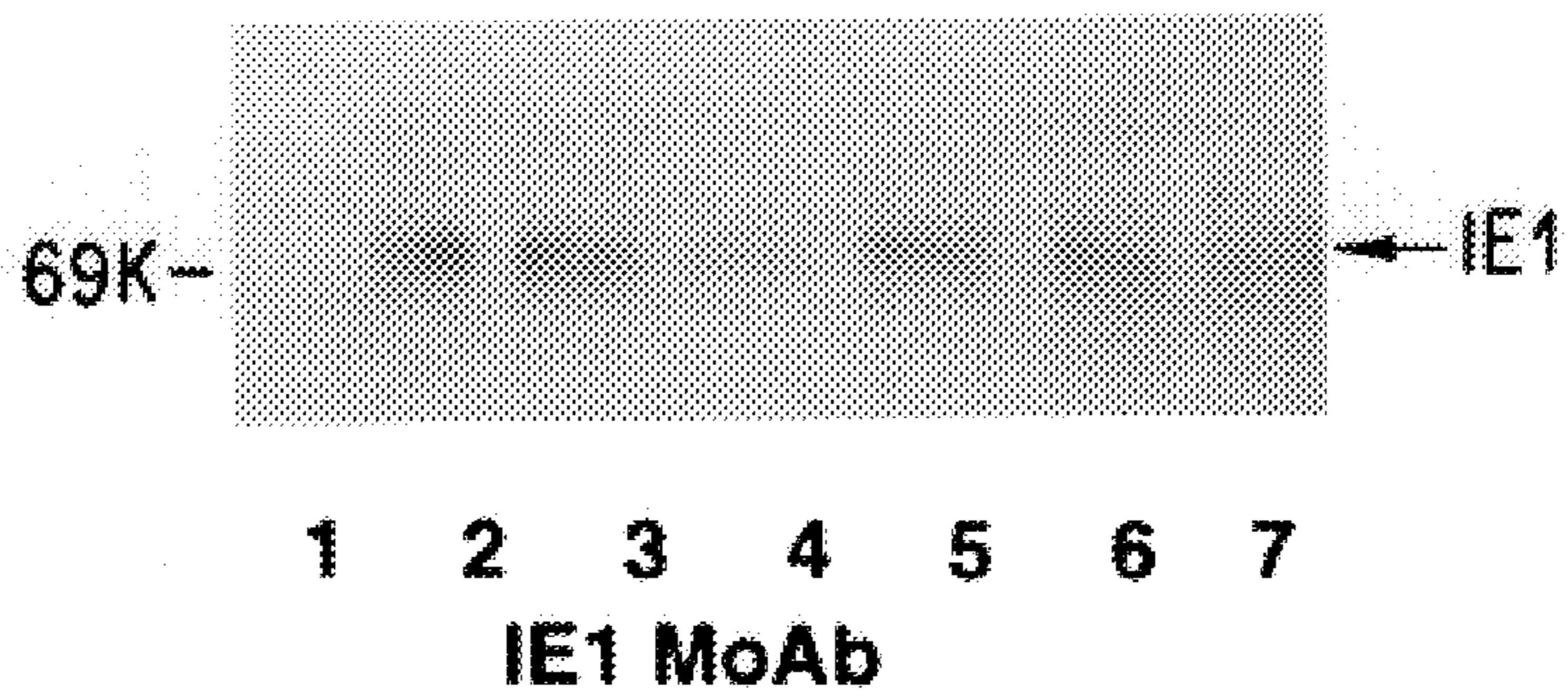


FIG. 9B

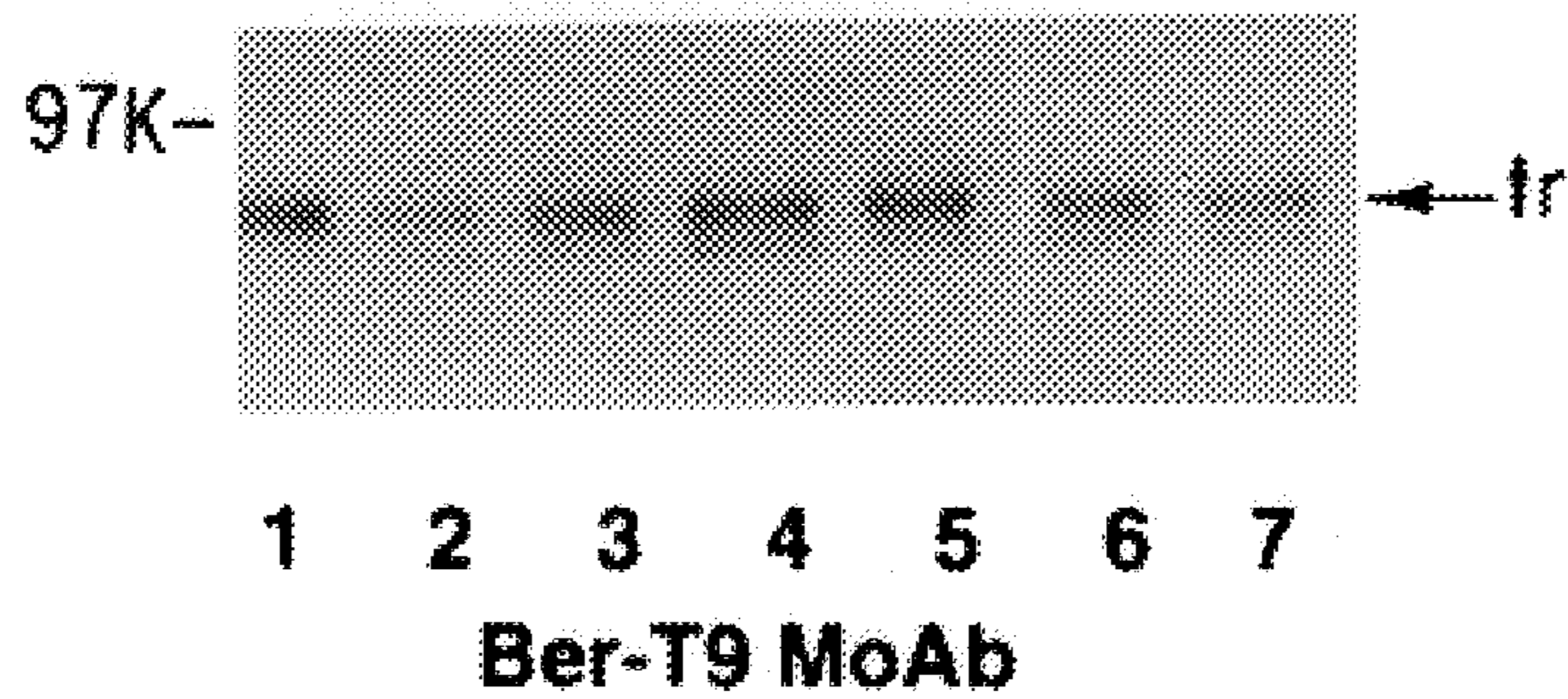


FIG. 9C

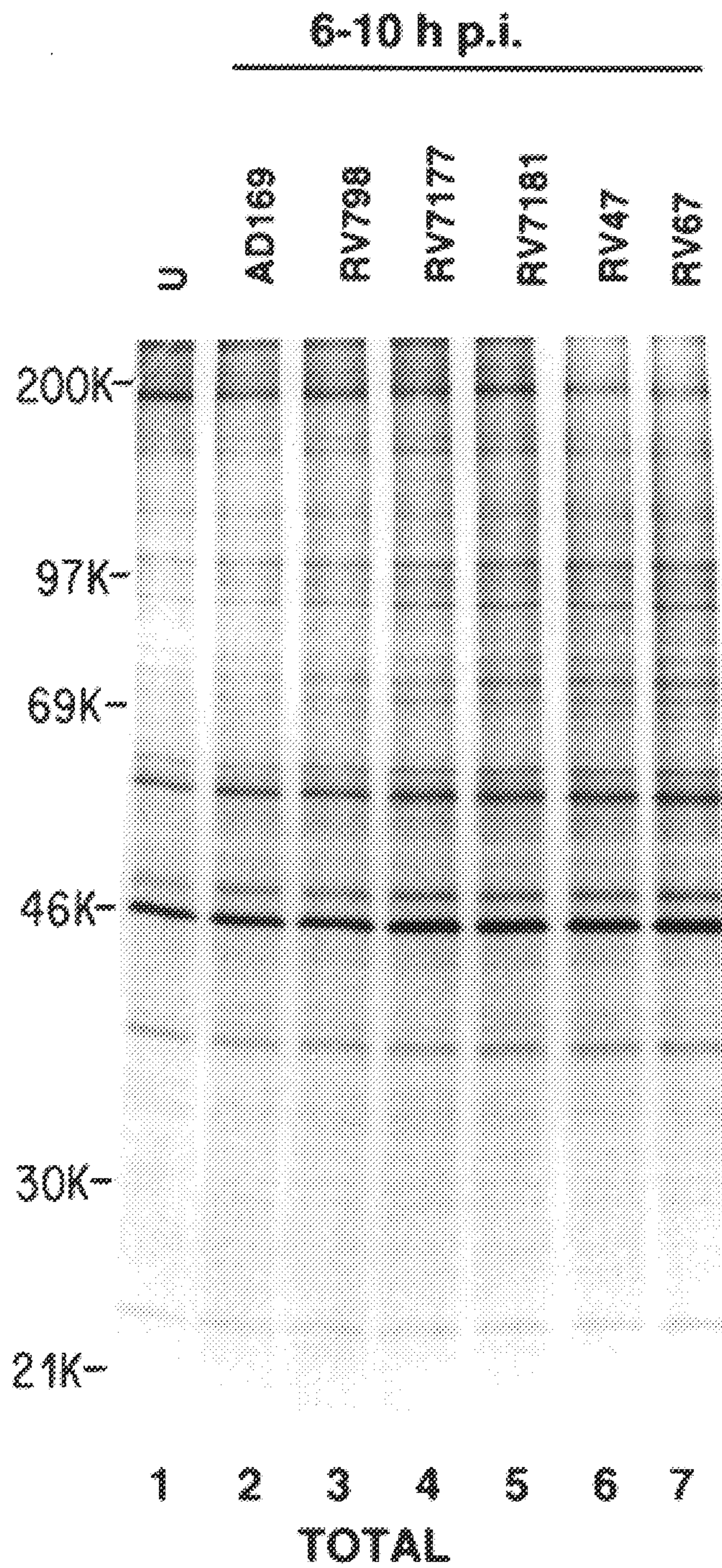
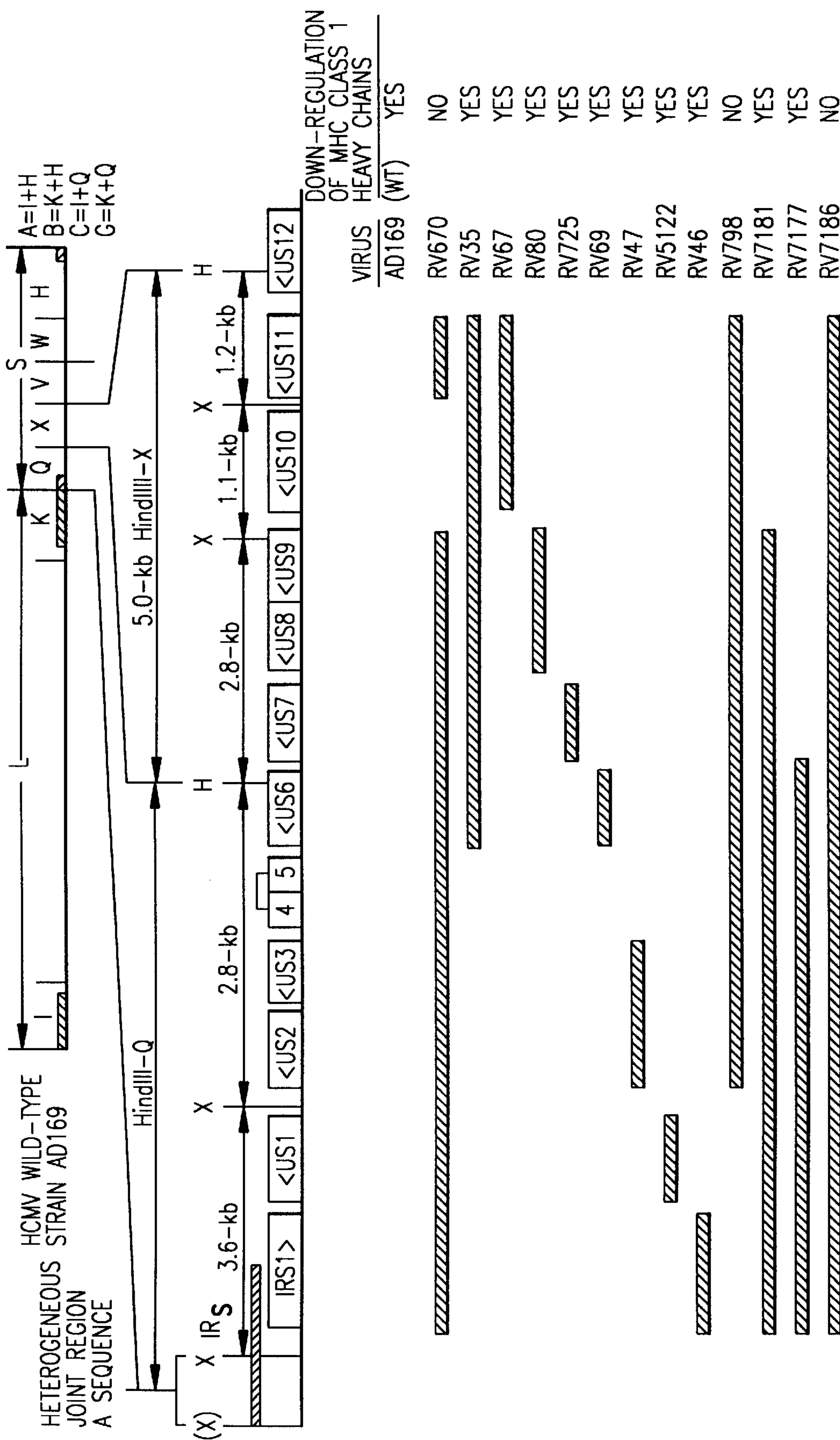


FIG. 9D



SUBREGION A

SUBREGION B

FIG.10

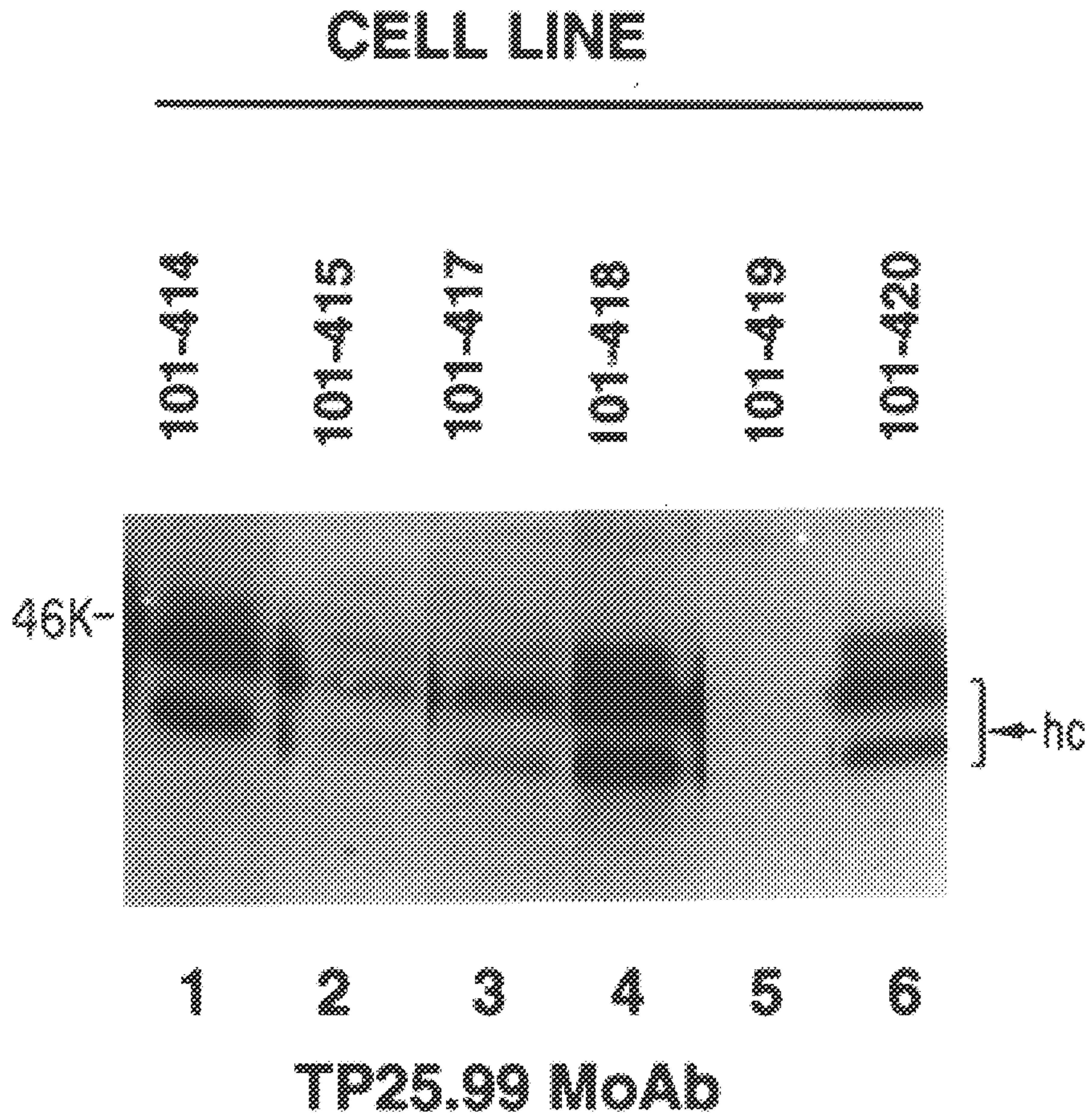


FIG. 11A

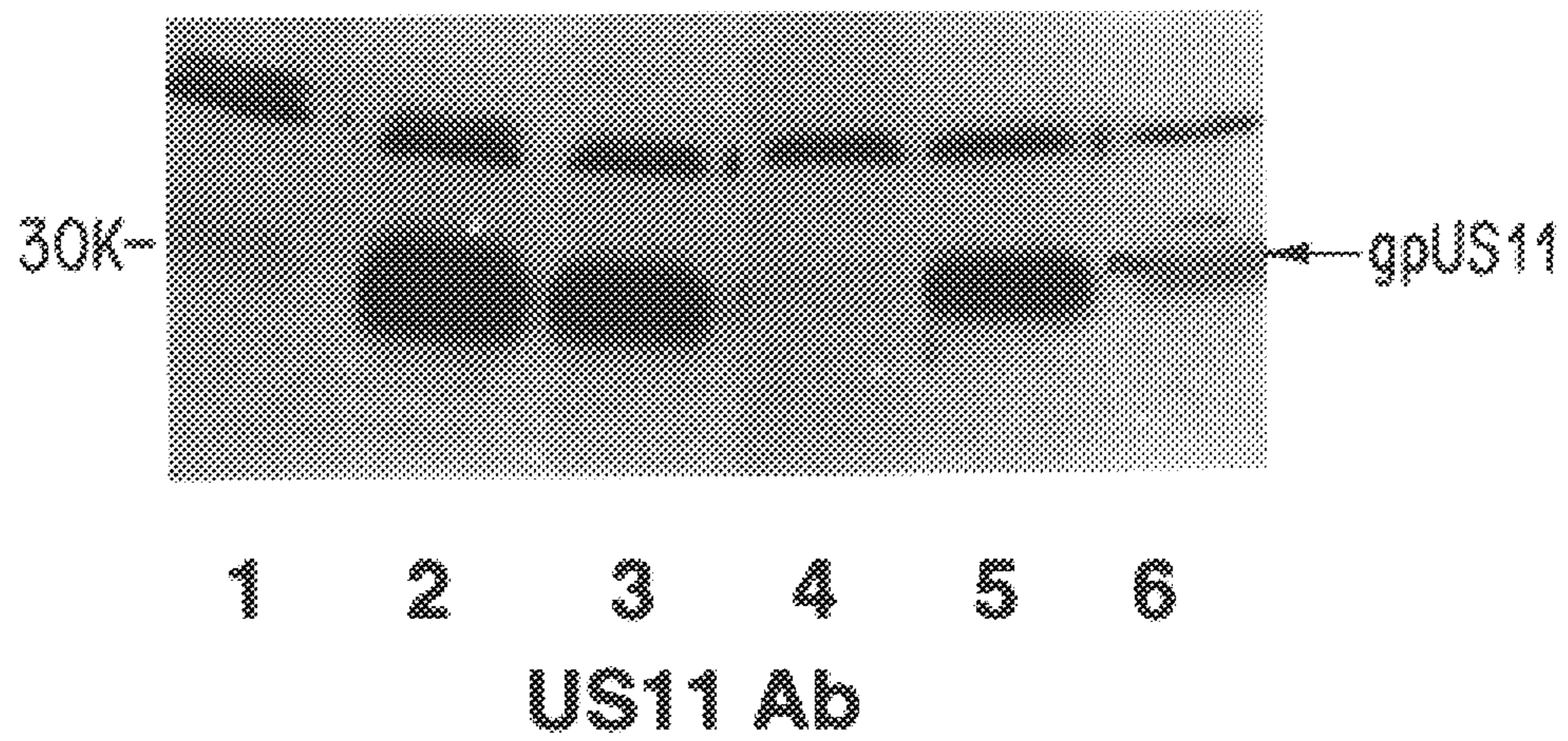


FIG. 11B

**IDENTIFICATION OF A HUMAN
CYTOMEGALOVIRUS GENE REGION
INVOLVED IN DOWN-REGULATION OF
MHC CLASS I HEAVY CHAIN EXPRESSION**

FIELD OF THE INVENTION

The present invention relates to recombinant mutant human cytomegalovirus (HCMV) which does not down regulate expression of cellular MHC class I heavy chains upon infection.

BACKGROUND OF THE INVENTION

Human cytomegalovirus (HCMV) is a betaherpesvirus which causes clinically serious disease in immunocompromised and immunosuppressed adults, as well as in some infants infected in utero or perinatally (Alford and Britt, 1990). The 230-kb dsDNA genome of HCMV was sequenced (Chee et al., 1990) and has at least 200 open reading frames (ORFs). For purposes of this application, open reading frame is defined as the portion of a gene which encodes a string of amino acids and hence may encode a protein. The function of some HCMV proteins are known or predicted due to their homology with other viral (esp. herpes simplex virus) and cellular proteins. However, for the majority of the HCMV ORFs, the function(s) of the proteins they encode is unknown.

In order to study HCMV gene function HCMV deletion mutants can be constructed in order to assess their in vitro growth properties (Jones et al., 1991; Jones and Muzithras, 1992). For purposes of this application deletion mutants are defined as human cytomegalovirus mutants which lack regions of the wild-type viral genome. This strategy involves site-directed replacement mutagenesis of selected HCMV gene(s) by a prokaryotic reporter gene, usually β -glucuronidase, although guanosine phosphoribosyltransferase can also be used. In this fashion, the recombinant virus can be isolated only if the replaced viral gene(s) is nonessential.

Several investigators have shown that infection by HCMV results in the down regulation of cellular MHC class I heavy chains (Browne et al., 1990; Beersma et al., 1993; Yamashita et al., 1993). For purposes of this application, down regulation is defined as reduction in either synthesis, stability or surface expression of MHC class I heavy chains. Such a phenomenon has been reported for some other DNA viruses, including adenovirus, murine cytomegalovirus, and herpes simplex virus (Anderson et al., 1985; Burget and Kvist, 1985; del Val et al., 1989; Campbell et al., 1992; Campbell and Slater, 1994; York et al., 1994). In the adenovirus and herpes simplex virus systems, the product of a viral gene which is dispensable for replication in vitro is sufficient to cause down regulation of MHC class I heavy chains (Anderson et al., 1985; Burget and Kvist, 1985). The gene(s) involved in class I heavy chain down regulation by murine cytomegalovirus have not yet been identified.

SUMMARY OF THE INVENTION

The present invention is directed to a recombinant mutant human cytomegalovirus which does not down regulate expression of cellular MHC class I heavy chains upon infection. Mutants RV 798 and RV 799 both deleted of open reading frames US2-US11, lose the ability to down regulate MHC class I heavy chains.

The present invention is also directed to a method to produce the recombinant mutant human cytomegalovirus

and a vaccine which utilizes the cytomegalovirus. One skilled in the art will use live attenuated HCMV vaccine lacking this gene region in order to elicit a better immune response, than one containing this gene region, based on the lack of class I down registration by the former. Therefore a virus lacking the region is a superior immunogen.

In addition, the HCMV gene involved in the MHC class I heavy chain down regulation can be incorporated into adenovirus vectors or similar virus based gene therapy vectors to minimize the immune response which will allow the use of the recombinant adenovirus or similar virus based gene therapy vectors to be used in gene therapy.

The invention may be more fully understood by reference to the following drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1E Detection of cell surface MHC class I by immunofluorescence-flow cytometry in HCMV-infected cells. Human foreskin fibroblast (HFF) cells were infected with the indicated virus at a multiplicity of infection of 5 PFU/cell for 72 h. At that time, cells were fixed in 1% paraformaldehyde and stained with primary antibody specific for HLA-A, B, C (W6/32) or control mouse IgG (isotype matched) followed by secondary FITC-conjugated goat anti-mouse IgG. Percent positive cells (5×10^3 total) and mean fluorescent intensity (MFI) were calculated on the basis of forward angle light scatter versus log-integrated 90° light scatter using the Immuno Program, Coulter MDADS I.

FIGS. 2A-2C Expression of MHC class I heavy chains in HCMV wild-type strain AD169-infected cells. FIG. 2A is a Western blot analysis. HFF cells were uninfected (U) or infected at a multiplicity of infection of 5 PFU/cell. At 24, 48, and 72 h postinfection total cellular proteins were harvested, electrophoresed through a 15% SDS-polyacrylamide gel, electroblotted to nitrocellulose, and probed with TP25.99 murine monoclonal antibody (specific for a non-conformational epitope on MHC class I heavy chains) using an ECL chemiluminescent detection kit (Amersham). FIGS. 2B and C are immunoprecipitation analyses. HFF cells were uninfected or infected (as above), either in the absence or presence (+PFA) of phosphonofomate and radiolabeled either for 4 h at late times postinfection (69-73 h) FIG. 2B or for 2 h at the indicated time postinfection FIG. 2C. Proteins were harvested immediately after radiolabeling and class I heavy chains were immunoprecipitated using TP25.99 murine monoclonal antibody.

FIGS. 3A-3J Organization of recombinant virus genomes. FIG. 3A, the; first line, is a schematic of the overall organization of the HCMV wild-type genome. Unique region sequences are shown by a line, while repeated region sequences are indicated by shaded boxes. Relevant HindIII fragments, within the L and S components, are indicated by letter designation (Oram et al., 1982). The second line is an expansion of the wild-type HindIII-Q, -X, and -V regions of the S component. The significant open reading frames, and their orientation, are shown as open boxes (Chee et al., 1990). The position of the IRs repeated sequences is indicated by the shaded rectangle. The locations of HindIII (H) and XhoI (X) restriction endonuclease sites are shown. FIGS. 3B-I show the genomic organization of the indicated HCMV mutant. In each case, the first line is the organization of the AD169 wild-type genome, the second line represents the organization of relevant sequences of the linearized plasmid used to make the recombinant virus. The slanted lines indicate the boundaries of the viral flanking sequences which may be involved in homologous recombination.

nation to create the desired mutation. The region deleted is indicated by a shaded box below the first line. FIG. 3J shows the derivation and organization of RV799. The first two lines are the same representations as FIGS. 3B–I, and the third line represents the organization of the relevant sequences of the linearized plasmid used to make RV799 from the RV134 parent (second line).

FIGS. 4A–4C show the analysis of heavy chain expression in cells infected with HCMV mutants. HFF cells were uninfected (U) or infected with the indicated virus (multiplicity of infection of 5 PFU/cell) and radiolabeled for 4 h at late times postinfection (69–73 h). Proteins were harvested immediately after radiolabeling. FIG. 4A is a radiograph of Class I heavy chains which were immunoprecipitated using TP25.99 murine monoclonal antibody. FIG. 4B is a radiograph of total radiolabeled proteins to verify approximately equivalent radiolabeling efficiency. FIG. 4C is a radiograph to verify equal progression through the viral replicative cycle. UL80 proteins were immunoprecipitated using anti-assembly protein rabbit polyclonal antiserum.

FIG. 5A–5C shows immunoprecipitation of class I heavy chains from RV798-, RV799-, RV134-, or AD169 wild-type-infected cells. HFF cells were uninfected (U) or infected with the indicated virus (multiplicity of infection of 5 PFU/cell) and radiolabeled for 2 h at late times postinfection (71–73 h). Proteins were harvested immediately after radiolabeling. FIG. 5A is a radiograph of Class I heavy chains which were immunoprecipitated using TP25.99 murine monoclonal antibody. Equivalent radiolabeling efficiency (FIG. 5B) and progression through the viral replicative cycle (FIG. 5C) were verified as described for FIG. 4B and 4C.

FIG. 6 is a radiograph showing the endoglycosidase H sensitivity of class I heavy chains synthesized in RV798-infected cells. HFF cells were infected with RV798 (multiplicity of infection of 5 PFU/cell) and radiolabeled for 2 h at early times (6–8 h) or late times (80–82 h) postinfection. For comparison purposes, uninfected cells were radiolabeled for 2 h. Proteins were harvested either immediately after radiolabeling (pulse) or after a 2 h chase (chase) in complete unlabeled media. Class I heavy chains were immunoprecipitated using TP25.99 murine monoclonal antibody. Immunoprecipitated protein were incubated for 6 h either in the presence (+) or absence (–) of 1.5 mU of endoglycosidase H, prior to SDS-polyacrylamide gel electrophoresis and fluorography. FIGS. 7A–7C show the immunoprecipitation of class I heavy chains from RV798-, RV7181-, RV7177-, or AD169 wild-type-infected cells. HFF cells were uninfected (U) or infected with the indicated virus (multiplicity of infection of 5 PFU/cell) and radiolabeled for 2 h at late times postinfection (65–67 h). Proteins were harvested immediately after radiolabeling. FIG. 7A is a radiograph of Class I heavy chains which were immunoprecipitated using TP25.99 murine monoclonal antibody. Equivalent radiolabeling efficiency (FIG. 7B) and progression through the viral replicative cycle (FIG. 7C) were verified as described for FIG. 4B–C.

FIGS. 8A–8D are photographs which show localization of US11 gene product (gpUS11) in infected cells by immunofluorescence. HFF cells were uninfected or infected with either AD169 wild-type or RV699 (deleted of the US11 gene) at a multiplicity of infection of 5 PFU/cell. After 8 h, uninfected and infected cells were fixed with 4% paraformaldehyde. Some cells were then permeabilized with 0.2% TRITON X-100™ (alkylanyl polyether alcohol). The primary antibody was rabbit polyclonal antisera raised against a US11 fusion protein (Jones and Muzithras, 1991). Fluorescence was visualized through a Zeiss microscope.

FIGS. 9A–9D show analysis of heavy chain expression in cells infected with HCMV mutants at early times postinfection. HFF cells were uninfected (U) or infected with the indicated virus (multiplicity of infection of 5 PFU/cell) and radiolabeled for 4 h from 6–10 h postinfection. Proteins were harvested immediately after radiolabeling. FIG. 9A is a radiograph of Class I heavy chains which were immunoprecipitated using TP25.99 murine monoclonal antibody. FIG. 9B is a radiograph in which, to verify approximately equal infection, the 72-kDa IE1 immediate-early protein was immunoprecipitated using the murine monoclonal antibody 9221. FIG. 9C is a radiograph of the immunoprecipitation of the cellular transferrin receptor with murine monoclonal antibody Ber-T9 to verify approximately equal expression of this glycoprotein. FIG. 9D is a radiograph of total radiolabeled proteins to verify approximately equivalent radiolabeling efficiency.

FIGS. 10 provides a summary of MHC class I heavy chain expression data from HFF cells infected with wild-type and mutant HCMV. In FIG. 10A, the first line is the overall organization of the HCMV wild-type genome, and the second line is an expansion of the wild-type HindIII-Q and -X regions of the S component. The ORFs are indicated by an unshaded rectangle; the unlabeled ORF overlapping US4 and US5 is US4.5 in FIG. 10(B), the deletions within the various HCMV mutants are indicated by the shaded rectangle. RV670 is deleted of IRS1–US9 and US11; RV35 is deleted of US6–US11; RV67 is deleted of US10–US11; RV80 is deleted of US8–US9; RV725 is deleted of US7; RV69 is deleted of US6; RV47 is deleted of US2–US3; RV5122 is deleted of US1; RV46 is deleted of IRS1; RV798 is deleted of US2–US11; RV7181 is deleted of IRS1–US9; RV7177 is deleted of IRS1–US6; and RV7186 is deleted of IRS1–US11. MHC class I heavy chain down regulation results are from immunoprecipitation experiments (using the heavy chain conformation-independent monoclonal antibody, TP25.99) in which HCMV-infected HFF cells were radiolabeled at late times postinfection. FIG. 10C shows location of the two subregions which contain gene(s) which are sufficient for MHC class I heavy chain down regulation. Subregion A contains ORFs US2–US5 (bases 193119–195607 Set forth in SEQ. ID NO:1) and subregion B contains ORFs US10 and US11 (bases 199083–200360 Set forth in SEQ. ID NO:2).

FIG. 11A–11B are western blots of cell lines expressing the HCMV US11 gene. Uninfected human U373-MG astrocytoma cells stably transformed with a US11 expression plasmid were analyzed by Western Blot analysis for MHC class I heavy chain expression (FIG. 11A) and for US11 expression (FIG. 11B) using the TP25.99 monoclonal antibody and the US11 polyclonal antisera, respectively.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

A recombinant HCMV mutant called RV670, has been constructed which expresses a marker gene (β -glucuronidase) in place of a group of viral genes. Upon infection of human fibroblast cells with this mutant, it is demonstrated that expression of the major histocompatibility complex (MHC) class I heavy chains are not reduced, as it is when wild-type HCMV infects these cells.

Unlike wild-type HCMV, the present invention's virus does not result in the down regulation of cellular MHC class I heavy chain protein expression. A 7kb region of the HCMV genome which contains genes which are required for down regulation of heavy chain expression is utilized in the invention.

One skilled in the art will appreciate that efficient antigen processing and presentation is required to activate and expand cytotoxic T-Lymphocyte precursors for an efficient cell mediated immune response. Efficient viral antigen presentation requires the continued expression of MHC class I proteins throughout infection. Infection of cells with RV670 results in continued expression of class I heavy chains.

One skilled in the art will appreciate that the claimed virus (RV670) or another human cytomegalovirus with a deletion of similar genes, can be utilized to produce an effective live vaccine since class I heavy chains are still expressed in RV670-infected cells, as they are in uninfected cells, and therefore viral antigen presentation, for the purpose of initiating a cytotoxic T cell response occurs.

In the present invention, flow cytometry and immunofluorescence experiments confirmed that cell surface expression of class I heavy chains are greatly reduced at late times postinfection in HCMV wild-type strain 8169 infected HFF cells. Radiolabeling-immunoprecipitation experiments indicates that down regulation of newly synthesized MHC class I heavy chains occurs throughout the course of infection, beginning at very early times (3 h) postinfection (FIG. 2C). This reduction has been reported to be at the post-translational level: class I heavy chains have a higher turnover rate in HCMV-infected cells than in uninfected cells (Beersma et al., 1993). Such instability of class I heavy chains results in a reduced cell mediated immune response to HCMV infection since viral peptides will be inefficiently presented. Thus, the reduction in class I heavy chain expression is important in terms of evasion of host's immune system in the establishment of persistent or latent infections by HCMV (Gooding, 1992).

We screened our bank of HCMV mutants which represent 18 ORFs which are dispensable for viral replication in tissue culture for their ability to cause down regulation of MHC class I heavy chains. A 7-kb region of the S component of the HCMV genome, containing ORFs US2-US11 (bases 193119-200360 Set forth in SEQ. ID NO:3), is clearly shown to contain genes which are required for this phenotype (data summarized in FIG. 10). Within this region, there are two subregions, each of which contain genes sufficient for heavy chain down regulation. Subregion A contains ORFs US2-US5 (bases 193119-195607 Set forth in SEQ. ID NO:1). It is proposed that US2 and US3 encode membrane glycoproteins (Chee et al., 1990). US3 is a differentially spliced gene which is expressed throughout the viral replicative cycle and encodes a protein with transcriptional transactivating function (Tenney and Colberg-Poley, 1991; Colberg-Poley et al., 1992; Tenney et al., 1993; Weston, 1988). Several smaller ORFs are also present in this subregion (between the ORFs US3 and US5), but their expression characteristics or functions have not been reported. Gretch and Stinski (1990) reported that there is a 1.0-kb early mRNA transcribed from this region of the HCMV genome, but it was not fine-mapped. It is not yet known which of these genes are involved in heavy chain down regulation.

Subregion B is also sufficient for MHC class I heavy chain reduction contains the US10 and US11 genes (FIG. 10), bases 199083-200360 Set forth in SEQ. ID NO:2. However, based on data using HCMV mutant RV670, which expresses wild-type levels of the US10 gene product (Jones et al., manuscript in preparation), US10 expression is not sufficient for down regulation of heavy chain expression (FIG. 2B). Thus, the genetic data implicates the US11 gene product as being required. We have demonstrated that US11 expression is sufficient to cause MHC class I heavy chain down regulation in stably transformed uninfected cells in the

absence of other MCVN proteins (FIG. 11). RNA and protein expression from both of these ORFs begins early and proceeds throughout the course of infection (Jones and Muzithras, 1991); US10 and US11 encode glycoproteins of 22-kDa (gpUS10) and 32-kDa, (gpUS11) respectively; both glycoproteins have N-linked sugar residues which are completely endoglycosidase H sensitive. These glycoproteins are retained in the endoplasmic reticulum or cis golgi. Consistent with this conclusion is the immunofluorescence data in which gpUS11 was not detected on the cell surface, but was detected in the cytoplasm of HCMV-infected cells (FIG. 8). The characteristics of HCMV gpUS11 (as well as gpUS10) are similar to the 25-kDa glycoprotein (E3-19K) encoded from the E3 region of adenovirus type 2. Ad E3-19K is nonessential for viral replication. It has been shown to contain endoglycosidase H-sensitive N-linked sugar residues, be retained in the endoplasmic reticulum, and bind MHC class I heavy chains; thereby preventing their transport to the cell surface (Anderson et al., 1985; Burgert and Kvist, 1985). In contrast to Ad E3-19K, a direct association between gpUS11 (or gpUS10) and class I heavy chains (i.e. by coimmunoprecipitation) was not detected (data not shown).

The identification of US2-US11 gene region as the region of the HCMV genome required for down regulation of MHC class I heavy chains is significant in several respects. As mentioned above, expression from this region of the genome throughout the course of infection acts to interfere with an effective cell mediated immune response. Surface expression of MHC class I molecules is required for antigen presentation to activate and expand cytotoxic T lymphocyte (CTL) precursors populations (Schwartz, 1985). In addition, they are further required for target recognition by the activated CTLs (Zinkernagel and Doherty, 1980). In MCMV, CTLs against the major immediate-early protein are protective against lethal infection by this virus (Jonjic et al., 1988). However, in HCMV infected individuals, the frequency of CTLs against the analogous HCMV immediate-early protein, IE1, are reported to be extremely rare (Gilbert et al., 1993). Recent studies have shown that IE peptides are more efficiently presented by interferon- γ -treated HCMV-infected cells, than by untreated infected cells (Gilbert et al., 1993). Interferon μ causes increased surface expression of MHC class I proteins. Thus, increasing the expression of class I heavy chains in HCMV-infected cells may be important in the efficient generation of IE-specific CTLs, or CTLs against other important HCMV antigens. A HCMV mutant deleted of the US2-US11 gene region would have this effect since the class I heavy chains are not down regulated when cells are infected with this mutant. Therefore, a deletion of this region of the viral genome is important in the development of an live HCMV vaccine to induce an effective anti-HCMV immune response.

Several years ago it was reported that the HCMV UL18 ORF encoded a protein which resembled MHC class I heavy chains (Beck and Barrell, 1988). It was hypothesized that the down regulation of heavy chains in HCMV-infected cells was due to competition of the UL18 gene product for β 2-microglobulin, which effectively prevented the normal association of class I heavy chains and β 2-microglobulin (Browne et al., 1990). This hypothesis was essentially dispelled when a HCMV mutant deleted of UL18 retained its ability to down regulate heavy chain expression (Browne et al., 1992). It remained possible that the UL18 gene product was only one of several HCMV genes whose expression is sufficient for this phenotype. However, the present invention data indicates that only genes within the US2-US11 region are sufficient for class I heavy chain down regulation.

The existence of two independent mechanisms which result in down regulation of MHC class I expression emphasizes the importance of this phenotype for successful infection and persistence in the host. One mechanism may serve as a backup system for the other, but also plausible is that there is cell type specificity for each system. In the case of the HFF cell system, both mechanisms are functional. However, in U373-MG cells, down regulation of heavy chain expression is more dependent on the presence of the subregion A. In that case, there may be qualitative or quantitative differences in cellular proteins which interact with subregion B gene products. A similar situation exists in the herpes simplex virus system. It was recently reported that the 88 amino acid US12 gene product (ICP47) is sufficient for class I heavy chain sequestering in the endoplasmic reticulum (York et al., 1994). However, expression of heavy chains is not affected in herpes simplex virus-infected mouse cells, although ICP47 is expressed in those cells and murine heavy chains are down regulated when expressed in an HSV-infected human fibroblast system (York et al., 1994).

A pharmaceutical composition may be prepared containing the recombinant HCMV mutant of the present invention in which the genome is devoid of a gene sequence capable of down regulating MHC Class I expression in infected cells. A stabilizer or other appropriate vehicle may be utilized in the pharmaceutical composition.

As discussed earlier, the recombinant HCMV mutant of the present invention which is devoid of the gene sequence capable of down regulating MHC Class I expression may be used in a vaccine for the prevention of cytomegalovirus infections. The vaccine comprises an effective amount of the recombinant HCMV mutant in a pharmaceutically acceptable vehicle. An adjuvant may be optionally added to the vaccine.

A method of immunizing an individual against cytomegalovirus may be carried out by administering to the individual an immunogenic amount of the recombinant HCMV mutant of the present invention which is devoid of the gene sequence capable of down regulating MHC Class I expression.

A method of preventing or reducing susceptibility in an individual to acute cytomegalovirus may be carried out by administering to the individual an immunogenic amount of the recombinant HCMV mutant of the present invention which is devoid of the gene sequence capable of down regulating MHC Class I expression.

Down regulation of MHC Class I expression in a cytomegalovirus infected cell may be controlled by a method having the steps of identifying a gene sequence capable of down regulating the major histocompatibility complex and deleting the identified gene sequence from the cytomegalovirus genome.

As discussed earlier, the gene sequence involved in the MHC Class I heavy chain down regulation can be incorporated into adenovirus vectors or similar virus based gene therapy vectors to minimize the immune response and allow the use of the vectors in gene therapy. One virus based gene therapy vector comprises the gene sequence of the open reading frame of US11. Another virus based gene therapy vector comprises the gene sequences of subregions A and B (open reading frames US2-US5 and US10-US11, respectively).

EXAMPLE 1

Virus and Cells. HCNV strain AD169 is obtained from the American Type Culture Collection and propagated accord-

ing to standard protocols known by those skilled in the art. Human foreskin fibroblast (HFF) cells were isolated in this laboratory and used below passage twenty (Jones and Muzithras, 1991). They were grown in Dulbeccos modified Eagle medium (DMEM) containing 10% fetal bovine serum and 25 mM HEPES.

DNA sequence. The numbering system of Chee et al. (1990) of the HCMV strain AD169 DNA sequence (Genbank accession number X17403) is used in the present invention.

Plasmids. Plasmids used for creation of HCMV mutants are constructed using the method described previously (Jones et al., 1991; Jones and Muzithras; 1992). Generally, the β -glucuronidase reporter gene is surrounded on each side by 1.5-kb of HCMV sequences which flank the gene(s) to be deleted from the virus. In each case, the plasmid DNA is linearized with a restriction enzyme which cuts within the prokaryotic backbone prior to transfection. The HCMV strain AD169 genomic DNA fragments are derived from either pHind-G, pHind-X, or pXba-P which contain the HindIII-G (bases 176844 to 195837 Set forth in SEQ. ID NO:4), -X (bases 195837 to 200856 Set forth in SEQ. ID NO:5), and XbaI-P (bases 200391 to 206314 Set forth in SEQ. ID NO:6) DNA fragments, respectively (Oram et al., 1982; Jones et al., 1991). pUS7/US3 contains the 1.7-kb PstI-PstI HCMV fragment (bases 194741 to 196447 Set forth in SEQ. ID NO:7; in pIBI30 vector [International Biotechnologies, Inc.]) derived from pHind-G and pHind-X.

To replace HCMV ORFs US11 through IRS1 by β -glucuronidase (i.e. RV7186; FIG. 3), pBgdUS11/IRS1 are constructed. Sequentially, this plasmid contains the 1.8-kb fragment PstI-XbaI fragment (bases 200391 to 202207 Set forth in SEQ. ID NO:8; containing US13, US12, and US11 promoter sequences; from pXba-P), β -glucuronidase, a 288-b SV40 fragment containing the early and late polyadenylation signals (from pRcCMV [Invitrogen]), and the 1.7-kb NcoI-NcoI fragment (bases 188062 to 189763 Set forth in SEQ. ID NO:9; containing J11 to IRL1 sequences; from pHind-G).

To replace HCMV ORFs US11 through US2 by β -glucuronidase (i.e. RV798; FIG. 3), pBgdUS11/US2 are constructed. Sequentially, this plasmid contains the 1.8-kb fragment PstI-XbaI fragment (bases 200391 to 202207 Set forth in SEQ. ID NO:8; containing US13, US12, and US11 promoter sequences; from pXba-P), β -glucuronidase, a 255-b fragment containing the US10 polyadenylation signal (bases 199021 to 199276 Set forth in SEQ. ID NO:10; from pHind-X), and the 1.3-kb NheI-ApaI fragment (bases 192033 to 193360 Set forth in SEQ. ID NO:11; containing C-terminal US2 to IRS1 sequences; from pHind-G).

To replace HCMV ORFs US11 through US6 by β -glucuronidase (i.e. RV35; FIG. 3), pBgdUS11/US6 was constructed. Sequentially, this plasmid contains the 1.8-kb PstI-XbaI fragment (bases 200391 to 202207 Set forth in SEQ. ID NO:8; containing US13, US12, and US11 promoter sequences; from pXba-P), β -glucuronidase, and the 1.5-kb HpaI-SstII fragment (bases 194062 to 195589 Set forth in SEQ. ID NO:12; containing C-terminal US6 to US3 sequences; from pHind-G). Replacement of HCMV ORFs US11-US10, or ORF US11 (singly), by β -glucuronidase (i.e. RV67 and RV699, respectively) were described previously (Jones et al., 1991).

To replace HCMV ORFs US9 through IRS1 by β -glucuronidase (i.e. RV7181; FIG. 3), pBgdUS9/IRS1 was constructed. Sequentially, this plasmid contains the 1.1-kb SalI-ApaI fragment (bases 199021 to 200171 Set forth in

SEQ. ID NO:13), the 351-b SV40 early promoter (from pRcCMV), β -glucuronidase, the 288-b SV40 polyadenylation signal fragment, and the 1.7-kb NcoI-NcoI fragment (bases 188062 to 189763 Set forth in SEQ. ID NO:9; containing J11 to IRL1 sequences; from pHind-G).

To replace HCMV ORFs US6 through IRS1 by β -glucuronidase (i.e. RV7177; FIG. 3), pBgdUS6/IRS1 was constructed. Sequentially, this plasmid contains the 1.7-kb NcoI-NcoI fragment (bases 188062 to 189763 Set forth in SEQ. ID NO:9; containing IRL1, J11, and IRS1 promoter sequences; from pHind-G), β -glucuronidase, the 255-b fragment containing the US10 polyadenylation signal (bases 199021 to 199276 Set forth in SEQ. ID NO:10; from pHind-X), and the 1.8-kb BsmI-SauI fragment (bases 196222 to 198030 Set forth in SEQ. ID NO:14; containing US7 to C-terminal US9 sequences; from pHind-X).

To replace HCMV ORFs US3 and US2 by β -glucuronidase (i.e. RV47; FIG. 3), pBgdUS3/US2 was constructed. Sequentially, this plasmid contains the 1.7-kb PstI-PstI fragment (bases 194741 to 196447 Set forth in SEQ. ID NO:7), a 180-b SmaI-HaeIII fragment containing the HSV-1 gH promoter (McKnight, 1980), β -glucuronidase, the 255-b US10 polyadenylation signal fragment, and the 1.3-kb NheI-ApaI fragment (bases 192033 to 193360 Set forth in SEQ. ID NO:11; containing C-terminal US2 to IRS1 sequences; from pHind-G).

To replace HCMV ORF US1 by β -glucuronidase (i.e. RV5122; FIG. 3), pBgdUS1 was constructed. Sequentially, this plasmid contains the 1.8-kb AatII-SstI fragment (bases 190884 to 192648 Set forth in SEQ. ID NO:15; containing IRS1 and US1 C-terminal sequences; from pHind-G), a 180-b SmaI-HaeIII fragment containing the HSV-1 gH promoter (McKnight, 1980), β -glucuronidase, the 255-b US10 polyadenylation signal fragment, and the 1.6-kb SphI-SphI fragment (bases 192934 to 194544 Set forth in SEQ. ID NO:16; containing US2 and C-terminal US3 sequences; from pHind-G).

To replace HCMV ORF IRS1 by β -glucuronidase (i.e. RV46; FIG. 3), pBgdIRS1 was constructed. Sequentially, this plasmid contains the 1.7-kb NcoI-NcoI fragment (bases 188062 to 189763 Set forth in SEQ. ID NO:9; containing IRL1, J11, and IRS1 promoter sequences; from pHind-G), β -glucuronidase, the 255-b fragment containing the US10 polyadenylation signal (bases 199021 to 199276 Set forth in SEQ. ID NO:10; from pHind-X), and the 1.2-kb NarI-XhoI fragment (bases 191830 to 193003 Set forth in SEQ. ID NO:17; containing C-terminal IRS1 and US1 sequences; from pHind-G). To delete HCMV ORFs US11 through US2 without insertion of a reporter gene (i.e. RV799; FIG. 3), pdUS11/US2 was constructed. Sequentially, this plasmid contains the 1.8-kb fragment PstI-XbaI fragment (bases 200391 to 202207 Set forth in SEQ. ID NO:8; containing US13, US12, and US11 promoter sequences; from pXba-P), β -glucuronidase, 65-b NruI-ApaI fragment containing the US10 polyadenylation signal (bases 199021 to 199086 Set forth in SEQ. ID NO:18; from pHind-X), and the 1.3-kb NheI-ApaI fragment (bases 192033 to 193360 Set forth in SEQ. ID NO:11; containing C-terminal US2 to IRS1 sequences; from pHind-G).

Isolation of recombinant mutant HCMV

Creation and isolation of recombinant mutant HCMV is done as described previously (Jones et al., 1991; Jones and Muzithras, 1992). HFF cells are split so that they are 70–80% confluent on the day of transfection. The cells are trypsinized and suspended to 5.6×10^6 cells per ml in DMEM/10% FCS/25 mM HEPES. The DNA is transfected using a modified calcium phosphate co-precipitation tech-

nique. 1.5 μ g of infectious HCMV DNA and 2.5 μ g of linearized plasmid DNA are mixed in the calcium chloride solution (300 μ l containing 10 mM Tris pH 7.0/250 mM calcium chloride) and chilled on ice. To initiate the co-precipitation, the DNA is removed from the ice and 300 μ l $2 \times$ HeBS pH 6.95 (at room temperature; $1 \times$ HeBS is 19.2 mM HEPES, 137 mM NaCl, 5 mM KCl, 0.8 mM sodium phosphate, 0.1% dextrose) is added dropwise with gentle mixing. After 1.5 min, the precipitate is placed on ice (to prevent further precipitate from forming). The precipitate is mixed with 3×10^6 cells (in suspension) and placed in a 82mm tissue culture plate. After 6 h at 37° C., the media is removed and the cells are shocked with 20% DMSO in $1 \times$ HeBS for 2 min. The cells are washed twice with PBS and growth media is added. The media is changed every 4–7 days. After 14 days, viral plaques are observed and the cells are overlaid with 0.5% agarose in DMEM containing 150 μ g/ml X-gluc (5-bromo 4-chloro 3-indol 1-glucuronide; Biosynth). Blue plaques (i.e. β -glucuronidase-positive mutant virus plaques) are picked several days after adding the overlay. Recombinant viruses were plaque purified three times. HCMV mutant RV799 is β -glucuronidase-negative and is isolated using a modification of the above procedure. In this case, β -glucuronidase-positive HCMV mutant RV134 is the parent virus (Jones et al., 1991). Thus, RV134 genomic DNA is used instead of wild-type strain AD169 DNA in the transfections. Primary plaques appearing on the primary transfection plates are picked at random and replated on HFF cells. After 10 days, the media is removed and the infected cells are overlaid with X-gluc-containing agarose as described above. In this case, white plaques (β -glucuronidase-negative mutant virus plaques) are picked 4 days later and plaque purified. The proper genomic organization of each of HCMV mutants is verified by DNA blot hybridization analysis as described previously (Jones et al., 1991).

Antibodies

Rabbit polyclonal antisera reactive with HCMV US11 proteins and HCMV UL80 proteins are described previously (Jones et al., 1991; 1994). Murine monoclonal antibodies W6/32, specific for a conformation-dependent epitope on the heavy chain of human MHC class I proteins, and Ber-T9, specific for the human transferrin receptor, are purchased. Murine monoclonal antibody TP25.99 (D'Urso et al., 1991), specific for a conformation-independent epitope on the heavy chain of human MHC class I proteins, is obtained from Dr. S. Ferrone (Department of Microbiology, New York Medical College, Valhalla, N.Y.). Murine monoclonal antibody 9221, specific for the HCMV IE1 protein, is purchased from Dupont.

Radiolabeling and immunoprecipitation of infected cell proteins

Pulse-chase radiolabeling is done according to standard protocol (Sambrook et al., 1989). HCMV-infected HFF cells (multiplicity of infection equals five) is pulse-labeled with 200 μ Ci of [35 S] methionine and [35 S] cysteine (NEN-DuPont) per ml in methionine/cysteine-free Dulbecco's modified Eagle medium (DMEM) at the indicated time period postinfection. The radioactive media is removed, the cells washed twice in complete DMEM, and chases are done for the indicated time in complete DMEM. Proteins are extracted using triple detergent lysis buffer (Sambrook et al., 1989). The cleared protein extracts (supernatant after centrifugation for 5 min at $15000 \times g$ and 4° C.) are retained for immunoprecipitation according to standard protocol (Sambrook et al., 1989). Proteins binding to antibodies are pelleted using protein A sepharose (Pharmacia). For immu-

noprecipitations of the human transferrin receptor, rabbit anti-mouse IgG (Pierce) are added prior to protein A sepharose. The washed immunoprecipitates were boiled in the presence of 2-mercaptoethanol and electrophoresed in denaturing polyacrylamide gels. The gels are fixed and soaked in 1M sodium salicylate fluor (Sambrook et al., 1989) prior to drying and autoradiography.

Immunofluorescence. Immunofluorescence assays are done according to standard protocol (Harlow, 1989). All procedures are done in 60mm tissue culture plates. Briefly, infected or uninfected HFF cells were fixed with 4% paraformaldehyde and permeabilized with 0.2% TRITON X-100™ (where indicated). After adding 3% bovine serum albumin in phosphate-buffered saline, the cells are held overnight at 4° C. The cells are treated sequentially with the following antisera, each for 30 min at room temperature: 10% HCMV-negative human serum (to block any Fc receptors); the indicated primary antibody; and FITC-conjugated anti-mouse or anti-rabbit IgG, as appropriate.

EXAMPLE 2

Class I down regulation in HCMV wild-type-infected human fibroblasts. We sought to ascertain the timing and nature of MHC class I heavy chain down regulation in the present invention's human foreskin fibroblast (HFF) cell culture system. By flow cytometry, HCMV strain AD169 wild-type-infected HFF cells are significantly reduced in the expression of class I heavy chains on their cell surface at late times postinfection (i.e. 72 h) using the conformation-dependent class I monoclonal antibody W6/32 (FIG. 1). In western analyses using the conformation-independent class I monoclonal antibody (TP25.99), it is demonstrated that the steady state level of class I protein is also reduced at late times postinfection (FIG. 2A). Because viral peptides are presented at the cell surface by class I complexes assembled after infection, we sought to assess the status of class I proteins synthesized at various times postinfection by immunoprecipitation of metabolically radiolabeled proteins. As shown in FIG. 2B, reduction in expression of class I heavy chains is detected both in the presence and absence of the viral DNA synthesis inhibitor, phosphonoformate. This indicates that viral immediate-early or early gene functions are sufficient for heavy chain reduction. In addition, it is demonstrated that heavy chain down regulation was detected at very early times postinfection: 3 h (FIG. 2C). Since this effect is observed using the conformation-independent antibody, the reduction reflects overall levels of newly synthesized heavy chains.

Screening of HCMV mutants for the loss of MHC class I down regulation

Several previously constructed HCMV deletion mutants, representing 18 nonessential ORFs (UL33, UL81, IRS1, US1–US13, US27–US28, and TRS1), are screened for heavy chain expression by flow cytometry and immunoprecipitation analyses. Only RV670, a mutant deleted of a 9-kb region within the S component of the HCMV genome (Jones and Muzithras, 1992), does not retain the wild-type down regulation phenotype (FIG. 4A). This mutant is deleted of at least 11 ORFs, IRS1 through US11 (except for US10), which includes the US6 family of genes (US6–US11) which putatively encode glycoproteins (Chee et al., 1990). To confirm this observation, two additional independently derived mutants which have the same deletion as RV670 and a new mutant, RV7186, deleted of the entire IRS1–US11 region (FIG. 3) are tested. Each is phenotypically identical to RV670 and stably expressed class I heavy chains. Previously, we constructed HCMV mutants deleted of US6

family ORFs, either individually or in groups (Jones and Muzithras, 1992), and similar deletion mutants within the adjacent IRS1-US3 region. By immunoprecipitation using the conformation-independent antibody, all of these mutants are shown to retain the ability to down regulate class I heavy chains (FIG. 4A) at late times postinfection in HFF cells. Control experiments indicate that radiolabeling is equivalent between the different infected cell cultures (FIG. 4B) and that infection proceeded to late times equally, as judged by pp65 (FIG. 4B) and UL80 protein (FIG. 4C) expression. These data indicate: (i) that more than one viral gene is sufficient for the reduction in class I heavy chains; or (ii) gene(s) between US3 and US6, deleted in RV670 and RV7186 but not the other mutants, is required for the phenotype.

Identification of a 7-kb region of the HCMV genome required for MHC class I down regulation

To further localize the region containing gene(s) involved in MHC class I heavy chain down regulation, additional HCMV replacement mutants containing deletions of multiple genes within the IRS1–US11 gene region are created (FIG. 3). One of these mutants, RV798, is deleted of genes from US2–US11. In HFF cells infected by RV798 and analyzed at late times postinfection, MHC class I heavy chains are not down regulated as they are in wild-type strain AD169-infected cells (FIG. 4A); in fact, a slight stimulation is observed. Several independently-derived deletion mutants identical to RV798 were examined similarly: all lacked the ability to down regulate class I heavy chains. To further confirm that the 7-kb HCMV US2–US11 region contains the gene(s) required for heavy chain down regulation, mutant RV799 is constructed which has the identical US2–US11 deletion as RV798, but is created by a different strategy. RV798 is derived from wild-type strain AD169 by inserting a β -glucuronidase marker gene in the place of US2–US11. In contrast, the parent of RV799 is RV134, a mutant which is β -glucuronidase-positive since it has a β -glucuronidase expression cassette inserted within the US9–US10 intergenic region (Jones et al., 1991). To create RV799, a plasmid is designed which upon recombination with the RV134 genome would simultaneously delete US2–US11 and the β -glucuronidase expression cassette (FIG. 3). The proper RV799 HCMV mutant is isolated as a white plaque in the presence of the β -glucuronidase substrate, since it is β -glucuronidase-negative. RV799, but not the RV134 parent, is phenotypically identical to RV798 (FIG. 5) Thus, since RV798 and RV799 are created by different strategies using parents which retained the ability to down regulate MHC class I heavy chains, this confirms that the gene(s) required for the phenotype are located within the 7-kb US2–US11 region (bases 193119–200360 Set forth in SEQ. ID NO:3).

To determine whether the proper surface expression of class I heavy chains occurred at late times postinfection with either RV798 or RV799, immunofluorescence assays are done. Using either the conformation-dependent (W6/32) or conformation-independent (TP25.99) monoclonal antibodies, surface expression of MHC class I heavy chains is detected in uninfected and RV798- and RV799-infected HFF cells, but not wild-type AD169-infected HFF cells. Proper maturation of class I heavy chains in uninfected cells yields endoglycosidase H resistant molecules. In contrast, class I heavy chains synthesized in AD169-infected cells are reported to be entirely endoglycosidase H sensitive (Beersma et al., 1993). As shown in FIG. 6, class I heavy chains synthesized in RV798-infected HFF cells, either at early or late times postinfection, are converted to the mature

endoglycosidase H-resistant form at a rate similar to those synthesized in uninfected cells. Taken together, these data indicate that MHC class I synthesis, processing, and surface expression are not impaired in cells infected with these HCMV mutants. Furthermore, the results indicate that the

7-kb region containing US2–US11 genes contain one or more genes required for heavy chain down regulation by HCMV.

Two subregions within the US2–US11 gene region contain genes which are involved in class I heavy chain down regulation

The region of the HCMV genome deleted in RV35 is from US6–US11, and US2–US11 in RV798 (FIG. 3). In RV35-infected HFF cells, MHC class I heavy chains are down regulated, but in RV798-infected cells they are not (FIG. 4A). This data indicates that one or more genes involved in heavy chain down regulation maps within the 2-kb subregion from ORF US2 through US5 (subregion A; bases 193119–195607 Set forth in SEQ. ID NO:7). To determine if this 2-kb subregion is required for class I heavy chain down regulation, HCMV replacement mutants RV7181 and RV7177 are examined. HCMV ORFs IRS1–US9 and IRS1–US6 are deleted, respectively, in these mutants; hence, subregion A is absent from both mutants. Experiments in infected HFF cells at late times postinfection indicates that both mutants retained the ability to efficiently down regulate class I heavy gene expression (FIG. 7). Therefore, when present in the HCMV genome, gene(s) within subregion A are sufficient for reduction of MHC expression (e.g. RV35), although their presence is not required for the phenotype. Furthermore, the cumulative data indicate that there are no HCMV genes within the identified 7-kb US2–US11 region (i.e. the region deleted in RV798) which are absolutely required for efficient heavy chain down regulation in infected HFF cells, suggesting that gene(s) from another portion of the US2–US11 gene region are also sufficient for the phenotype at late times postinfection.

Evidence indicating that the US11 gene product is involved in MHC class I heavy chain down regulation

In HFF cells infected with mutant RV7181, deleted from IRS1–US9 (FIG. 3), MHC class I heavy chain expression is down regulated, in contrast to RV798-infected HFF cells (FIG. 7). This data suggests that a second subregion (subregion B), comprised of the US10 and US11 genes (bases 199083–200360 Set forth in SEQ. ID NO:2), is involved in reduction of heavy chain expression. However, the expression of US10 from the context of the HCMV genome is not sufficient for heavy chain down regulation. HCMV mutant RV670 expresses US10 at steady-state levels similar to wild-type and is deleted of all of the other ORFs in the 7-kb US2–US11 gene region, but it does not cause down regulation of MHC class I heavy chains in infected HFF cells (FIGS. 2B and 4A).

US11 encodes a 32-kDa glycoprotein (gpUS11) containing N-linked, but not O-linked, carbohydrates which are completely sensitive to endoglycosidase H, indicating that the sugars are in the high mannose form. gpUS11 is detected throughout infection, beginning at very early times (i.e. 3 h) and continuing through late times postinfection. However, levels of gpUS11 in the infected cell are most abundant at approximately 8 h postinfection. To determine its location in the infected cell, rabbit polyclonal antisera (Jones and Muzithras, 1991) is used in immunofluorescence assays of wild-type strain AD169-infected cells. Uninfected and RV699-infected HFF cells are used as negative controls. RV699 is an HCMV mutant which is isogenic with AD169, except for a deletion of the US11 ORF (Jones et al., 1991).

In cells fixed and permeabilized at 8 h postinfection, cytoplasmic fluorescence which obscured definition of the nucleus is observed in AD169-infected HFF cells, but not in either negative control cells (FIG. 8). In general, the specific fluorescence is more intense in the perinuclear area. There is no specific fluorescence detected in non-permeabilized cells (FIG. 8). The fluorescence and endoglycosidase-H sensitivity data indicate that gpUS11 is not a cell surface glycoprotein. From the translated DNA sequence, gpUS11 is predicted to have hydrophobic domain near its N- and C-termini (Weston and Barrell, 1986) which are putative signal sequence and transmembrane domain, respectively. Thus, gpUS11 is associated with intracytoplasmic membranes, possibly the endoplasmic reticulum.

Down regulation of MHC class I expression at early times postinfection by HCMV mutants

MHC class I expression in wild-type strain AD169-infected cells are shown to begin at very early times postinfection (FIG. 2C). To determine if any of the mutants are deficient for this early down regulation, immunoprecipitation experiments are performed using extracts from infected HFF cells radiolabeled from 6–10 h postinfection. The level of class I heavy chains are reduced during this early period postinfection in HFF cells with each of the mutants, except for RV798, the mutant deleted of the entire 7-kb US2–US11 region (FIG. 9A). Control experiments demonstrated that the different mutant-infected cells are equally infected and radiolabeled (FIG. 9B and D). Expression of another cellular glycoprotein, the transferrin receptor, is not differentially affected by the various mutants (FIG. 9C). Thus, genes required for heavy chain down regulation at early times postinfection are the same as those necessary for reduction at late times postinfection. Moreover, expression of gene(s) from either subregion identified to be involved in down regulation of heavy chain expression at late times postinfection are sufficient for reduction at very early times postinfection.

EXAMPLE 3

Recombinant HCMV (RV798) Vaccine Preparation

HCMV vaccines are prepared in using a method described previously (Elek and Stern, 1974). HCMV mutant RV798 is grown on MRC-5 human diploid lung fibroblasts (CCL171 [American Type Culture Collection]) or human foreskin fibroblasts (MRHF [BioWhittaker]). Cells are infected at a multiplicity of infection equal to one in Dulbecco's modified Eagle medium (DMEM) containing 5% calf serum and 5% fetal calf serum. After 24 h, the medium is removed and the cells washed three times with either Hank's balanced salt solution or Dulbecco's phosphate-buffered saline. Fresh DMEM medium without serum is added; the infected cells are incubated 4 days after the appearance of late viral cytopathic effect (usually 7 days postinfection). After a preclearing centrifugation step (6,000×gravity for 20 min at 18° C.), cell-free virus is pelleted by centrifugation at 15,500×gravity for 1 h at 18° C. The pelleted virus is resuspended in Dulbecco's phosphate-buffered saline containing 25% sorbitol and stored in aliquots at –70° C. The titer of RV798 vaccine stock is determined using standard procedures on human foreskin fibroblasts (Wentwork and French, 1970). The vaccine is administered by subcutaneous inoculation of approximately 10³–10⁷ plaque forming units into the deltoid region of the upper arm, as described previously (Elek and Stern, 1974; Gehrz et al., 1980; Starr et al., 1981).

EXAMPLE 4

gpUS11 is sufficient for down regulation of MHC class I heavy chains

To determine if the US11 gene product, in the absence of any other viral gene products, is capable of causing heavy chain down regulation, the US11 coding region (bases 199716 to 200360 Set forth in SEQ. ID NO:19 [Chee et al., 1990]) and some non-coding flanking sequences, encompassing bases 199683 to 200391 Set forth in SEQ. ID NO:20, are cloned into a eukaryotic expression plasmid under the transcriptional control of the constitutive HCMV major immediate-early promoter-enhancer. Human U373-MG astrocytoma cells (HTB 17 [American Type Culture Collection]) are transfected with this plasmid (Sambrook et al., 1989) and stably transformed cells are selected in the presence of 0.375 μ g/ml of puromycin, since the plasmid also encodes for the prokaryotic puromycin resistance gene. Clones are picked and expanded into cell lines. Those expressing gpUS11 are identified by western blot analysis; different cell lines expressed varying amounts of US11. MHC class I heavy chain expression in these cell lines is analyzed in a similar fashion. As shown in FIG. 11, expression of US11 is inversely correlated with the expression of class I heavy chains. These data prove that expression of HCMV US11 is sufficient for the down regulation of MHC class I heavy chain expression, in the absence of any other viral gene products.

REFERENCES

- Alford, C. A., and W. J. Britt. 1990. Cytomegalovirus, p. 1981–2010. In D. M. Knipe and B. N. Fields (ed.), *Virology*, 2nd ed. Raven press, New York.
- Anderson, M., S. Paabo, T. Nilsson, and P. A. Peterson. 1985. Impaired intracellular transport of class I MHC antigens as a possible means for adenoviruses to evade immune surveillance. *Cell* 43:215–222.
- Beck, S., and B. G. Barrell. 1988. Human cytomegalovirus encodes a glycoprotein homologous to MHC class I antigens. *Nature* 331:269–272.
- Beersma, M. F. C., M. J. E. Bijlmakers, and H. L. Ploegh. 1993. Human cytomegalovirus down regulates HLA class I expression by reducing the stability of class I H chains. *J. Immunol.* 151:4455–4464.
- Browne, H., M. Churcher, and T. Minson. 1992. Construction and characterization of a human cytomegalovirus mutant with the UL18 (class I homolog) gene deleted. *J. Virol.* 66:6784–6787.
- Browne, H., G. Smith, S. Beck, and T. Minson. 1990. A complex between the MHC class I homolog encoded by human cytomegalovirus and β 2 microglobulin. *Nature* 347:770–772.
- Burgert, H. G., and S. Kvist. 1985. An adenovirus type 2 glycoprotein blocks cell surface expression of human histocompatibility class I antigens. *Cell* 41:987–997.
- Campbell, A. E., J. S. Slater. 1994. Down-regulation of major histocompatibility complex class I synthesis by murine cytomegalovirus early gene expression. *J. Virol.* 68:1805–1811.
- Campbell, A. E., J. S. Slater, V. J. Cavanaugh, and R. M. Stenberg. 1992. An early event in murine cytomegalovirus replication inhibits presentation of cellular antigens to cytotoxic T lymphocytes. *J. Virol.* 66:3011–3017.
- Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchinson, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. Weston, and B. G. Barrell. 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr. Top. Microbiol. Immunol.* 154:125–169.

- Colberg-Poley, A. M., L. D. Santomenna, P. P. Harlow, P. A. Benfield, and D. J. Tenney. 1992. Human cytomegalovirus US3 and UL36–38 immediate-early proteins regulate gene expression. *J. Virol.* 66:95–105.
- del Val, M., K. Munch, M. Reddehase, and U. Koszinowski. 1989. Presentation of CMV immediate-early antigen to cytotoxic T lymphocytes is selectively prevented by viral genes expressed in the early phase. *Cell* 58:305–315.
- D'Urso, C. M., Z. Wang, Y. Cao, R. Tatake, R. A. Zeff, and S. Ferrone. 1991. Lack of HLA class I antigen expression by cultured melanoma cells FO-1 due to a defect in β 2m gene expression. *J. Clin. Invest.* 87:284–292.
- Elek, S. D., and H. Stern. 1974. Development of a vaccine against mental retardation caused by cytomegalovirus infection in utero. *Lancet* 1:1–5.
- Gehrz, R. C., W. R. Christianson, K. M. Linner, K. E. Groth, and H. H. Balfour, Jr. 1980. Cytomegalovirus vaccine: specific humoral and cellular responses in human volunteers. *Arch. intern. Med.* 140:936–939.
- Gilbert, M. J., S. R. Riddell, C-R. Li, and P. D. Greenberg. 1993. Selective interference with class I major histocompatibility complex presentation of the major immediate-early protein following infection with human cytomegalovirus. *J. Virol.* 67:3461–3469.
- Gooding, L. R. 1992. Virus proteins that counteract host immune defenses. *Cell* 71:5–7.
- Gretch, D. R., and M. F. Stinski. 1990. Transcription of the human cytomegalovirus glycoprotein gene family in the short unique component of the viral genome. *Virology* 174:522–532.
- Jones, T. R., and Muzithras, V. P. 1991. Fine mapping of transcripts expressed from the US6 gene family of human cytomegalovirus strain AD169. *J. Virol.* 65:2024–2036.
- Jones, T. R., and V. P. Muzithras. 1992. A cluster of dispensable genes within the human cytomegalovirus genome short component: IRS1, US1 through US5, and the US6 family. *J. Virol.* 66:2541–2546.
- Jones, T. R., V. P. Muzithras, and Y. Gluzman. 1991. Replacement mutagenesis of the human cytomegalovirus genome: US10 and US11 gene products are nonessential. *J. Virol.* 65:5860–5872.
- Jones, T. R., L. Sun, G. A. Beberitz, V. P. Muzithras, H-J. Kim, S. H. Johnston, and E. Z. Baum. 1994. Proteolytic activity of human cytomegalovirus UL80 protease cleavage site mutants. *J. Virol.* 68: 3742–3752.
- Jonjic, S., M. de Val, G. M. Keil, M. J. Reddehase, and U. Koszinowski. 1988. A nonstructural viral protein expressed by a recombinant vaccinia virus protects against lethal cytomegalovirus infection. *J. Virol.* 62:1653–1658.
- Mavromara-Nazos, P., M. Ackerman, and B. Roizman. 1986. Construction and properties of a viable herpes simplex virus 1 lacking coding sequences of the alpha 47 gene. *J. Virol* 60:807–812.
- McKnight, S. L. 1980. The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. *Nucl. Acids Res.* 8:5949–5964.
- Oram, J. D., R. G. Downing, A. Akrigg, A. A. Dollery, C. J. Duggleby, G. W. G. Wilkinson, and P. J. Greenaway. 1982. Use of recombinant plasmids to investigate the structure of the human cytomegalovirus genome. *J. Gen. Virol.* 59:111–129.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Schwartz, R. H. 1985. T lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. *Ann. Rev. Immunol.* 3:237–261.

- Starr, S. E., J. P. Glazer, H. M. Friedman, J. D. Farquhar, and S. A. Plotkin. 1981. Specific cellular and humoral immunity after immunization with live Towne strain cytomegalovirus vaccine. *J. Infect. Dis.* 143:585–589.
- Tenney, D. J., and A. M. Colberg-Poley. 1991. Human 5 cytomegalovirus UL36–38 and US3 immediate-early genes: temporally regulated expression of nuclear, cytoplasmic, and polysome-associated transcripts during infection. *J. Virol.* 65:6724–6734.
- Tenney, D. J., L. D. Santomenna, K. B. Goudie, and A. M. 10 Colberg-Poley. 1993. The human cytomegalovirus US3 immediate-early protein lacking the putative transmembrane domain regulates gene expression. *Nucl. Acids Res.* 21:2931–2937.
- Wentworth, B. B., and French, L. 1979. Plaque assay of 15 cytomegalovirus strains of human origin. *Proc. Soc. Exp. Biol., Med.* 135:253–258.
- Weston, K. 1988. An enhancer element in the short unique region of human cytomegalovirus regulates the produc-

- tion of a group of abundant immediate early transcripts. *Virology* 162:406–416.
- Weston, K., and B. G. Barrell. 1986. Sequence of the short unique region, short repeats, and parts of the long repeats of human cytomegalovirus. *J. Mol. Biol.* 192:177–208.
- Yamashita, Y., K. Shimokata, S. Mizuno, H. Yamaguchi, and Y. Nishiyama. 1993. Down-regulation of the surface expression of class I MHC antigens by human cytomegalovirus. *Virology* 193:727–736.
- York, I. A., C. Roop, D. W. Andrews, S. R. Riddell, F. L. Graham, and D. C. Johnson. 1994. A cytosolic herpes simplex virus protein inhibits antigen presentation to CD8+ T lymphocytes. *Cell* 77:525–535.
- Zinkernagel, R. M., and P. C. Doherty. 1980. MHC restricted cytotoxic T cells: studies on the biological role of polymorphic major transplantation antigens determining T cell restriction specificity. *Adv. Immunol.* 27:51–177.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 20

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2489 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

GCACACGAAA AACCGCATCC ACATCATAGA CAAGTTACAG TCCACAGTCA CATACACGAT      6 0
AAACAATAACC AACAGGGTAA TGTTTATGGA GTAAAACACT ATTGTCCAGG CCACATGCGT      1 2 0
GTATGACTTC CGCACCATCC CGTACTGCAT GTTCCACATG TACGCGCTAG ACGTGTAATC      1 8 0
CACTCGCAGT TCGGGGACGC AACGCAGCCA GATCACATCC CCTTGCAGTA CCAGACGCAG      2 4 0
GGCTAGCGTC TCGAAGATCG GCATCACATC TAAGTTCCGC ACGTTCCACT TTAACGACTC      3 0 0
CCCGGGAACG AACTCCACGT CGTCGGCGTG TACGTACAGG TTCTCTCCCA CGCCGCCATA      3 6 0
ATCGGCCTTC GGATCGAAGA CGAACCGACT CATGTTGCCC ACGATGCTCC CCCGAGCAAA      4 2 0
CAACTTGCCG TTGTCAATGT AGCACCGGTT GTCCTCGATT TGAAACCAGG GATGCTTGGC      4 8 0
CGTGGACTTC CAGGGCCGGA GCGCGTCTTC CCCGGCTTTA GTGATTCCAT CGGGCAGGCG      5 4 0
GATCAAGGGA CCCATGGAGG TCCAAAGACC CACCCAGGCT TTCCAGAGAT TGTTCATGGT      6 0 0
GAAACAGCGT GTGGACTGTA CGCTCTTTCC CAATTTATAT CCCAGAGTAG TGACGTGAGC      6 6 0
CCAGCCACCT CCCAGATTCC TGACGTTTTG GTTGTCTTTC CTGCCAATTC CTCCCGTAAA      7 2 0
CTTATGATTA TCCTAGCCCA TTCCCGATAA AAATACACGG AGACAGTAGA TAGAGTTACG      7 8 0
AATAAACCGG TTTATTTATT CAAGTGTCTC AGGAGATTAT TGAACGAGCG TGGATACCAC      8 4 0
GCCGTCGTCA GTTCATGGTG GCATTGAGCA GCCATAGCAC CAGAGTCCCG GCGCCCGGTA      9 0 0
TCAGACACGC TGACCTACCG GCGCCTTCG AGTCCGTACC CCGCGGCCTG GGTGTTAGAG      9 6 0
TCCGTACCTT GCAGCCAGG TAGGTTTCAG GTACCAGCTG GTTCGTACCT GTTAAATAAA     1 0 2 0
TCGCAGACGG GCGCTCACCC CTACGGTCAG GAGCACAAGA ACAACCAGAG AGAACAGATA     1 0 8 0

```

-continued

TACGAGCAGG	GTTCTGAACA	GCAGACCCCA	ATTGTCGTCT	CTCATGCTTC	GCTGAAGGTA	1140
CCAGTTGATG	GTCTGAGAGC	TATAGTCCAT	CCTCACCTGA	GGAACACACG	CGGCATATTT	1200
CTTGGGGTCT	CCCCACCTCG	TAGACAACGT	GATGTCCACC	ATATCCACGG	TGTGCGTCAC	1260
CGGGTGCCCA	CCGATGTTCC	ACTCGAAATA	GGCTCCGCGC	TCATCATGGT	GGTACTGCTC	1320
ACCGGACACC	TGCAGTCTGT	CCATGTAAGA	TTGAGAGACG	ATACCCACGT	TCACAAAGTG	1380
TTTCTCGGTG	AAGTTGCCCG	ACATCCTCCC	CTTGAAGTAC	AGCATGCCCA	TATGGAACCA	1440
GCATTGGTTC	TCCTCCTACTC	GAAAGTGGGC	CGATCTGATC	TCCGATACCA	CCACATCCAG	1500
GGGCCGGGGC	ACCGAGTCCG	CGAGTCTCAG	GAACAAGACG	GCCAGGATCG	CGAGCACCAA	1560
CACCGGCTTC	ATGGCTCCGA	AGGTCCGCTG	CTCGGCTCCG	CTCACCGCTC	CGGTCTGGCT	1620
GCAGCAGTGC	TTCGCTGAGA	AGTAGCGTGT	GGACTGAACG	GTGTTTTTTGA	ATATATAGCG	1680
TTTCTTGGTG	ACGTTGTTTC	CCCTACGTAG	TAGGCAACTA	CGTGCCAAAA	GAGGCGTTAC	1740
GGTACTTTCC	GTACTGGGAT	TTCCAAACCG	GGACTTTCCA	CACGGCGGTT	TCAACACCGG	1800
GACTTTTCAC	ACGGTGATTT	CGGCACCGGG	ACTTTCCGCA	CGGCGGTTTC	GCCACCGCTG	1860
ACGTTCTCAT	CGCCGCCCAC	GTCAACGGTG	GCGACACCGT	ACTTTCCCAT	GCGGTTTATA	1920
AACGTCAAGA	GTCACGTCAG	TCGCCACCCC	CCATTACACG	GCGATATCCC	GATAGGGCAT	1980
GAGGGGACCC	GGGTGTCGCG	ACATGTGCGC	GACAGGTGCG	GATTAGTGGT	CGTGTCGCGA	2040
CATGGACGTG	CAGGGGGATG	TCTGTGCGGA	TAGAGTTGAT	GTGACAGCCC	GCTACACCTC	2100
TCTGTGCGGA	CATGCATACA	CAACGGGCCG	GCTTGTGCGC	GATTGTGCGG	ACATATCGTT	2160
ATCAGTTAGC	GACCGGAGTT	GTCTATCGCG	ACATATCGTC	GACTATCGCG	ACAGAAAAAA	2220
TACCGTTCGT	AGAGAATGCC	GTGTTGAAGG	AACGCGCTTT	TATTGAGACG	ATAAAACAGC	2280
ATCAGGAGCC	ACAACGTCGA	ATCCCACGTC	CAGTCGATTC	GTATGTTATG	CTGCACAGCA	2340
ATGCTAGAAT	AACAACCAGC	AGGGTAATCC	CGCAACATAA	ATACAAAGTC	ACAGCGAAGA	2400
ATCCGTGTCG	TTCTATCAAG	CGAAACGCGT	TCCAAACGGC	CCCGTCACAG	ACGCAGTTAT	2460
TCATAAGCGT	TAACAACCGG	TGGCTAGGA				2489

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1278 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTCGCGAGGT	GGATAATAAC	CGCATATCAG	GAGGAGGGAT	CGGGTGATGA	CGCAGGCCCC	60
GCAGAACAGT	CCGAAATAAA	TTTTTAGTAT	TGCCCCATAG	TCGCCTAGAT	ACCAGAGGTA	120
CGTTAAGTTC	ATCAAAAACGC	CCATCGGGCGT	CCCGGAATCG	TATACCGGGC	ACACGAAGCG	180
TTCATAACAA	TCCCAGGAGG	CGAGTGTTAG	GGTAGCAGAG	TAGTTTCGGG	GTCGGTTTCC	240
TTCCGGCGAC	GACAGTTCCG	TGGGCAGCAG	AATGTACAGC	GCCTCGGTAG	CTGTGCGGGT	300
GCCTTCCACG	AGGATGGGCT	GCCGGTGCCT	TTCGTGATTT	TCCCCGTCGT	GTAGCCAAGC	360
CGAGGCCCGC	AAAGTCTTAG	GCGAGGGGAA	TTGTCCATAG	AGTTTCACCG	CACCCTTCAG	420
TACATGGTTC	TGAATAACAC	AGCCGCACGT	GAAGTAGGTA	GGTTCTCTCG	TCTCCTCCGT	480
GGCTGCCGCC	ACCACTCCCA	GCCACCACAA	CAGGCAGATC	GCCAGAGGGT	TCCGGAGGCT	540

-continued

TCCCCGGCGT	AGCATGGTTT	TGGGTAAAG	CAAAAAGTCT	GGTGAGTCGT	TTCCGAGCGA	6 0 0
CTCGAGATGC	ACTCCGCTTC	AGTCTATATA	TCACCACTGG	TCCGAAAACA	TCCAGGGAAA	6 6 0
ATGTCGGTGC	AGCCAACCTT	TCACATACAG	CCCCCAAAC	ACTTGAATCA	CTGCCACCAT	7 2 0
CATCAGCGTA	TACTGCGCCG	ACTTAATCGT	GAGCGCGTAG	TACGCCATTA	GACGGCGATC	7 8 0
TTCGAACAAT	AGTCGTTTGA	TGTCCTCTAA	CGAGCTCCAC	AGGGGAACCC	AAGGCACGAG	8 4 0
GCACCGGGGT	TCGCACTCTA	CATAATAAGT	TTGGCATTGG	TGGCAGGGGG	AAAAGTAGAA	9 0 0
CAACACGAGT	TTTGTGCGTT	GGGGAACACG	ATAGTCCCGG	AGCCAGTAGC	GTTTTGCGAC	9 6 0
GAGGCTTTTCG	GAGACGTCCT	CCACCGGCGT	CGGCACTCGA	TCCGCGTAGC	CCTCCAGCGT	1 0 2 0
CTGGTAGTAC	ACCCGGGGTG	TCGGCGTGGG	CACGGACAGG	TTCCCGCGCA	GGGTCCACAG	1 0 8 0
AGCCTCCAGT	CGACCGCCCG	ATCGGAGCAC	GCAGCGCGCC	TCGGAATACT	CTACTCGGTA	1 1 4 0
CTCCGAAAACA	TCGGACAGAG	GCGGTAACGG	CTCCGTCTCC	ACCAAGGGCG	GAGGTTTCATC	1 2 0 0
GAAAAGAGTC	AAGGATAATT	CAGGCATACT	ACCCGCGACC	GGGGCCAGA	GGGCTAGAAT	1 2 6 0
AAGCATTACA	AGGTTTCAT					1 2 7 8

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7242 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCACACGAAA	AACCGCATCC	ACATCATAGA	CAAGTTACAG	TCCACAGTCA	CATACACGAT	6 0
AAACAATACC	AACAGGGTAA	TGTTTATGGA	GTAAAACACT	ATTGTCCAGG	CCACATGCGT	1 2 0
GTATGACTTC	CGCACCATCC	CGTACTGCAT	GTTCCACATG	TACGCGCTAG	ACGTGTAATC	1 8 0
CACTCGCAGT	TCGGGGACGC	AACGCAGCCA	GATCACATCC	CCTTGCAGTA	CCAGACGCAG	2 4 0
GGCTAGCGTC	TCGAAGATCG	GCATCACATC	TAAGTTCCGC	ACGTTCCACT	TTAACGACTC	3 0 0
CCCGGGAACG	AACTCCACGT	CGTCGGCGTG	TACGTACAGG	TTCTCTCCCA	CGCCGCCATA	3 6 0
ATCGGCCTTC	GGATCGAAGA	CGAACCGACT	CATGTTGCC	ACGATGCTCC	CCCGAGCAA	4 2 0
CAACTTGCCG	TTGTCAATGT	AGCACCGGTT	GTCCTCGATT	TGAAACCAGG	GATGCTTGGC	4 8 0
CGTGGACTTC	CAGGGCCGGA	GCGCGTCTTC	CCCGGCTTTA	GTGATTCCAT	CGGGCAGGCG	5 4 0
GATCAAGGGA	CCCATGGAGG	TCCAAAGACC	CACCCAGGCT	TTCCAGAGAT	TGTTTCATGGT	6 0 0
GAAACAGCGT	GTGGACTGTA	CGCTCTTTCC	CAATTTATAT	CCCAGAGTAG	TGACGTGAGC	6 6 0
CCAGCCACCT	CCCAGATTCC	TGACGTTTTG	GTTGTCTTTC	CTGCCAATTC	CTCCCGTAAA	7 2 0
CTTATGATTA	TCCTAGCCCA	TTCCCGATAA	AAATACACGG	AGACAGTAGA	TAGAGTTACG	7 8 0
AATAAACCGG	TTTATTTATT	CAAGTGTCTC	AGGAGATTAT	TGAACGAGCG	TGGATACCAC	8 4 0
GCCGTCGTCA	GTTTCATGGT	GCATTGAGCA	GCCATAGCAC	CAGAGTCCCG	GCGCCCGGTA	9 0 0
TCAGACACGC	TGACCTACCG	GGCGCCTTCG	AGTCCGTACC	CCGCGGCCTG	GGTGTTAGAG	9 6 0
TCCGTACCTT	GCAGCCCAGG	TAGGTTTCAG	GTACCAGCTG	GTTTCGTACCT	GTTAAATAAA	1 0 2 0
TCGCAGACGG	GCGCTCACCC	CTACGGTCAG	GAGCACAAGA	ACAACCAGAG	AGAACAGATA	1 0 8 0
TACGAGCAGG	GTTCTGAACA	GCAGACCCCA	ATTGTCGTCT	CTCATGCTTC	GCTGAAGGTA	1 1 4 0
CCAGTTGATG	GTCTGAGAGC	TATAGTCCAT	CCTCACCTGA	GGAACACACG	CGGCATATTT	1 2 0 0
CTTGGGGTCT	CCCCACCTCG	TAGACAACGT	GATGTCCACC	ATATCCACGG	TGTGCGTCAC	1 2 6 0

-continued

CGGGTGCCCA	CCGATGTTCC	ACTCGAAATA	GGCTCCGCGC	TCATCATGGT	GGTACTGCTC	1 3 2 0
ACCGGACACC	TGCAGTCTGT	CCATGTAAGA	TTGAGAGACG	ATACCCACGT	TCACAAAGTG	1 3 8 0
TTTCTCGGTG	AAGTTGCCCG	ACATCCTCCC	CTTGAAGTAC	AGCATGCCCA	TATGGAACCA	1 4 4 0
GCATTGGTTC	TCCTCCACTC	GAAAGTGGGC	CGATCTGATC	TCCGATACCA	CCACATCCAG	1 5 0 0
GGGCCGGGGC	ACCGAGTCCG	CGAGTCTCAG	GAACAAGACG	GCCAGGATCG	CGAGCACCAA	1 5 6 0
CACCGGCTTC	ATGGCTCCGA	AGGTCCGCTG	CTCGGCTCCG	CTCACCGCTC	CGGTCTGGCT	1 6 2 0
GCAGCAGTGC	TTCGCTGAGA	AGTAGCGTGT	GGACTGAACG	GTGTTTTTTGA	ATATATAGCG	1 6 8 0
TTTCTTGGTG	ACGTTGTTTC	CCCTACGTAG	TAGGCAACTA	CGTGCCAAAA	GAGGCGTTAC	1 7 4 0
GGTACTTTCC	GTACTGGGAT	TTCCAAACCG	GGACTTTCCA	CACGGCGGTT	TCAACACCGG	1 8 0 0
GACTTTTCAC	ACGGTGATTT	CGGCACCGGG	ACTTTCCGCA	CGGCGGTTTC	GCCACCGCTG	1 8 6 0
ACGTTCTCAT	CGCCGCCAC	GTCAACGGTG	GCGACACCGT	ACTTTCCCAT	GCGGTTTATA	1 9 2 0
AACGTCAAGA	GTCACGTCAG	TCGCCACCCC	CCATTACACG	GCGATATCCC	GATAGGGCAT	1 9 8 0
GAGGGGACCC	GGGTGTCGCG	ACATGTCGAC	GACAGGTGCG	GATTAGTGGT	CGTGTCGCGA	2 0 4 0
CATGGACGTG	CAGGGGGATG	TCTGTCGCGA	TAGAGTTGAT	GTGACAGCCC	GCTACACCTC	2 1 0 0
TCTGTCGCGA	CATGCATACA	CAACGGGCCG	GCTTGTCGGC	GATTGTCGCG	ACATATCGTT	2 1 6 0
ATCAGTTAGC	GACCGGAGTT	GTCTATCGCG	ACATATCGTC	GACTATCGCG	ACAGAAAAAA	2 2 2 0
TACCGTTCGT	AGAGAATGCC	GTGTTGAAGG	AACGCGCTTT	TATTGAGACG	ATAAAACAGC	2 2 8 0
ATCAGGAGCC	ACAACGTCGA	ATCCCACGTC	CAGTCGATTC	GTATGTTATG	CTGCACAGCA	2 3 4 0
ATGCTAGAAT	AACAACCAGC	AGGGTAATCC	CGCAACATAA	ATACAAAGTC	ACAGCGAAGA	2 4 0 0
ATCCGTGTCG	TTCTATCAAG	CGAAACGCGT	TCCAAACGGC	CCCGTCACAG	ACGCAGTTAT	2 4 6 0
TCATAAGCGT	TAACAACCGG	TGGCTAGGAT	GAATATCCAA	ATCACAGGGC	AGTAGCCGAC	2 5 2 0
GGACTCGTTG	ACAGGTCAGC	CTACCCTCAA	GGTTCCTATC	GTTCCGGACGG	GATTTGTGCG	2 5 8 0
TTTTAGGCCT	CTTTTTCGCC	GCCTGCAAGC	ATTGGTGCGC	AAAGTCCTCA	CCCAGCTGTT	2 6 4 0
TCCAGCTATC	ATCTGCATCT	GTGCAGTCCC	CTGTATCGTT	GTAACAAACG	GGTCTGTGCG	2 7 0 0
ACTTCGTTCT	CGGAACACAA	GCTTGTTGTC	GCGGAGACAG	AGAGAGAAGG	GTTTTCGGGT	2 7 6 0
CACGCGAAGA	CCGCTCACCG	GGGGTCGGCA	ACGCACACAT	CAACAGAAAA	CCGAGACGAA	2 8 2 0
TCAAGAGATC	CATAGTGAAG	GAGTGATATC	GACGTGCTTA	CGAAACGGCG	ATTATATATG	2 8 8 0
TTCTCAACAA	TACCGCCCTA	CGTTGTATGA	TGTAACGTGT	GACGTGAGTC	TGATCCAACA	2 9 4 0
CTGAACGCTT	TCGTGCTGTT	TTTCATGCAG	CTTTTACAGA	CCATGACAAG	CCTGACGAGA	3 0 0 0
GCGTTCATCG	GGGCATGAAG	TACGCATTAC	ACAAACTCCA	TATATTTGTT	ACGATAGAAT	3 0 6 0
ACGGAACGGA	GGAGGCTTTC	GCCACACCTA	TCCTGAAAGC	GTTGCATTCT	TTATGATAGG	3 1 2 0
TGTGACGATG	TCTTTACCAT	TCCCACGGCT	GCTTTGCGTG	ATGATGACAT	TCATCATGTA	3 1 8 0
TTTCCATTCA	CACATACCTT	TTGTGCATAC	GGTTTATATA	TGACCATCCA	CGCTTATAAC	3 2 4 0
GAACCTAACA	GTTTATTAGC	CCTTGACAGG	ATAGGTCAAA	AGATTATATG	TAGGTTTTCC	3 3 0 0
GGTAAACCGA	ATTGTGATAT	TTCTCTGCAG	GAAATAGAAC	AGCCTGGTAC	CTATAAAACG	3 3 6 0
GACAATGCAG	TACTGTAGCA	GCGTAACCAA	GTAGGTCCAC	ATGAACACGT	ACAAAATTAT	3 4 2 0
GGTAAGCCAT	CGTTTTTCAT	ACCACAGCCT	GTAGCTGTCTG	TACATGAATG	AGGACGGTCG	3 4 8 0
AGGAACCCAG	GGTAGTTGTA	ATTGGGGGCG	ACATTCGTAC	TGTCCAGAAG	ACAATTGCAC	3 5 4 0
GGGTTTCAGT	GAGATGAGTA	CTTTAGCGAT	GTCGGCGGGG	GCGCTACGTT	TCACCGTGAC	3 6 0 0
GGTGAGAACT	TGACCGTCGT	TTTGTATTTT	ATGAGGCACG	TTATACAAGC	CACTGGTATC	3 6 6 0

-continued

ATGAAGGATG	ACCTCTGATG	CGATGTGAGG	ATTAAATTGT	CCCTCAAACC	GCCAAACGCT	3 7 2 0
GGTCATGTTT	CCACCGTCAA	TTACGCAGCT	GACGGTGTGA	GATACCACGA	TGTTGGACTT	3 7 8 0
AGGTTTGGGG	GCTAATTGCC	TTTTTACAAA	TTCCCTTCTG	TATTGCAGGT	CCTGCTGCCA	3 8 4 0
CTGCTTTTCC	GTGCGGAAAG	TCGCCATGTC	TTCCACACGT	GTGGCGACGA	TAGACGCCAC	3 9 0 0
CAAGGTAGCT	ACCAGAAGCA	GCTGGATCCG	CATGGCATT	CCGTATGTCA	ATTAGAAAGT	3 9 6 0
TGAGCGGACA	CGGTTATCGT	TCCTGGCGGA	TATAAGTATA	TAAACGCGAG	TTAGCCTTTC	4 0 2 0
CCGTCCGTTT	TGTACACCCG	TTCCCCACAC	AAATGACGAA	TACGACCTTT	TTTTTTATAA	4 0 8 0
AAATAAACCA	CGTGTATTAT	ATAAAAACAT	TTACATAGAA	AAGAGACACA	CGGATCAACA	4 1 4 0
TAAGGACTTT	TCACACTTTT	GGGGTACACA	GGCGTGCCAC	CGCAGATAGT	AAGCGCTGGA	4 2 0 0
TACACGGTAC	ACAGTCCTGG	CCAGCACGTA	TCCCAACAGC	AGCACCATCG	CCATACAGAT	4 2 6 0
GGCGATCACG	ACCCCGAGCT	CTAAGTGTCT	GTATTCATAG	TGTAGTCGCC	GCAGGTTATC	4 3 2 0
CACTGAATTC	CCGTAACTGA	AATAACGTAT	ATGGTACCGA	GGCTGGCACC	ACATGGGTTT	4 3 8 0
GCATTTGGTG	CACGGCACCA	AATGCAGAGT	GAGATGGTCC	AAGTCCGTGG	GCACCCACTG	4 4 4 0
GCGCAAACGG	AATACGGCTT	CGGTGGTCTC	CACGAGGCAC	TCCGGGGCGT	GCAGACGGCC	4 5 0 0
CCACTTTCGT	CCGCGACGGC	CCGACCAGCC	GACCCGAGCC	ACTATCCCTT	TCTCGGGATA	4 5 6 0
GAACGTACCC	TGTACACGCC	ACACAGCGTC	CAACACGCCG	TCCTTGACGA	CGCAGCTGGC	4 6 2 0
CTGATAGCTG	GACACGTTGT	TAAGCGGCGG	AAAGCGAAAC	TGACGTGCCG	GCGGAGCCAC	4 6 8 0
ATAGTTCGGT	TCACCGTGTT	GTCGCGGTTC	GTCCTCCCTA	TAGTAATAGT	AGTCGTCGTC	4 7 4 0
CTCATAGGGG	TTGCCGGCGT	GAGCCAGCGT	TACCCAACAG	CAGCCCAGGC	CGACGAGGAG	4 8 0 0
GCGCAGCCAC	CGCCTCATGG	CGGCTTCGCC	AGTCAATCGT	CTTTAGCCTC	TTCTTCCCGT	4 8 6 0
GAGGTCCTTC	CGGTGGCGCG	GTGCCGACCT	CGGACCCAGG	GACGTATCCA	CCTCAGGTAC	4 9 2 0
ACACAGCAGG	CTACCTGGAC	ACCGAAGCTG	AACAAGGCTA	CGTGTTTCAC	AAACTGCACC	4 9 8 0
AGTACCACAT	AGAGGAATGT	CAGGTAGCGT	CTCTCCGCAA	ACAGCCGTTC	CAAGTCTGAG	5 0 4 0
GGCGTTACCC	GCAGCGGCAA	CCAGGGCAGC	CTGGACGCCG	GCCGGCAATG	GAGCACGCTC	5 1 0 0
CGGTTACAGG	CACTGCAGGG	GTAAACGGTT	AACATCACGT	AAGAGAGTCG	TGCGTCCACC	5 1 6 0
TGTGGGAGCT	CAGTTTCGTA	ACGTAGAGCC	CCGTCATTTT	CCAGCTGGGG	TGCGCCGACC	5 2 2 0
TTGAAATGGG	TCGCGCTCCG	CTCGTTACCC	CAGGTGCCGT	AGGCTCTCGG	GGCCGTATCG	5 2 8 0
GAGAAGTTGC	CACGCACAAG	CCAGGCGGCC	ACGAGTACCC	CGTGCTGGAC	GTAACATTCG	5 3 4 0
GACACGGAAC	TGGAGACACG	GTAGCCGGAC	ACGTCCCCAA	ACCCGCGAGG	GTAAGTGGGC	5 4 0 0
AGACGGACGG	ACTTGCTATT	TGACAACGGA	CAGATACGAG	ACGACGAGGA	CGCAGACGAC	5 4 6 0
TCGTCGCTGG	ACCACGACAA	CCGGAGCGAC	TCCTTGGAGC	GGCTCGAGAG	TACACTTACT	5 5 2 0
GCGATCAGAC	ACCAGTGCCA	GAAGAAGGAA	CAGGTGGACG	GGGACCACAG	GATCATAGCC	5 5 8 0
GCCGGCACCG	CGGCCGGCCG	CAGGAAGCCG	CCCGGC	CGTCTGTGTG	CGGGAGCCGA	5 6 4 0
AACACCGTGC	CTCTTTATAT	CGTCCCGACG	TGACGCGAGT	ATTACGTGTC	AGGGGAAACC	5 7 0 0
CCCGTCACGA	CGAACGTGAT	TTGTAAGTGA	CGCGGGGTGC	TGACGGGGTT	CGGCCCGAGA	5 7 6 0
GGTGACGGAG	CGCCTCACGT	CAGTATGATG	TCCGATCCGC	GTCAGCCCCG	ACGTGGTTGT	5 8 2 0
GGTCACCGAA	ACCCACGTTT	ATATGGACGT	TGAGAGCAGC	GCCTGACCAC	ATGATTCATC	5 8 8 0
ATACCATTTT	TCGGAATCGG	GCCCATGCCG	GGAAAGCACA	TTCCTTTTCA	GTAAACAACA	5 9 4 0
ATGACATCAT	AACAAATCAT	TTTATTGCGG	AGGTGGATAA	TAACCGCATA	TCAGGAGGAG	6 0 0 0
GGATCGGGTG	ATGACGCAGG	CCCCGCAGAA	CAGTCCGAAA	TAAATTTTTA	GTATTGCCCC	6 0 6 0

-continued

ATAGTCGCCT	AGATAACCAGA	GGTACGT TAA	GTT CATCAA	ACGCCCATCG	GCGTCCCGGA	6 1 2 0
ATCGTATACC	GGGCACACGA	AGCGTTCATA	ACAATCCCGG	GAGGCGAGTG	TTAGGGTAGC	6 1 8 0
AGAGTAGTTT	CGGGGTCGGT	TTCCTTCCGG	CGACGACAGT	TCCGTGGGCA	GCAGAATGTA	6 2 4 0
CAGCGCCTCG	GTAGCTGTGC	CGGTGCCTTC	CACGAGGATG	GGCTGCCGGT	GCCTTTCGTG	6 3 0 0
ATTTTCCCCG	TCGTGTAGCC	AAGCCGAGGC	CCGCAAAGTC	TTAGGCGAGG	GGAATTGTCC	6 3 6 0
ATAGAGTTTC	ACCGCACCCCT	TCAGTACATG	GTTCTGAATA	ACACAGCCGC	ACGTGAAGTA	6 4 2 0
GGTAGGTTCT	CTCGTCTCCT	CCGTGGCTGC	CGCCACCACT	CCCAGCCACC	ACAACAGGCA	6 4 8 0
GATCGCCAGA	GGGTTCCGGA	GGCTTCCCCG	GCGTAGCATG	GTTTTGGGTT	AAAGCAAAAA	6 5 4 0
GTCTGGTGAG	TCGTTTCCGA	GCGACTCGAG	ATGCACTCCG	CTTCAGTCTA	TATATCACCA	6 6 0 0
CTGGTCCGAA	AACATCCAGG	GAAAATGTGC	GTGCAGCCAA	CCTTTCACAT	ACAGCCCCCA	6 6 6 0
AAACACTTGA	ATCACTGCCA	CCATCATCAG	CGTATACTGC	GCCGACTTAA	TCGTGAGCGC	6 7 2 0
GTAGTACGCC	ATTAGACGGC	GATCTTCGAA	CAATAGTCGT	TCGATGTCCT	CTAACGAGCT	6 7 8 0
CCACAGGGGA	ACCCAAGGCA	CGAGGCACCG	GGGTTTCGCAC	TCTACATAAT	AAGTTTGGCA	6 8 4 0
TTGGTGGCAG	GGGGAAAAGT	AGAACAACAC	GAGTTTTGTG	CGTTGGGGAA	CACGATAGTC	6 9 0 0
CCGGAGCCAG	TAGCGTTTTG	CGACGAGGCT	TTCGGAGACG	TCCTCCACCG	GCGTCGGCAC	6 9 6 0
TCGATCCGCG	TAGCCCTCCA	GCGTCTGGTA	GTACACCCGG	GGTGTCCGGC	TGGGCACGGA	7 0 2 0
CAGGTTCCCG	CGCAGGGTCC	ACAGAGCCTC	CAGTCGACCG	CCCGATCGGA	GCACGCAGCG	7 0 8 0
CGCCTCGGAA	TACTCTACTC	GGTACTCCGA	AACATCGGAC	AGAGGCGGTA	ACGGCTCCGT	7 1 4 0
CTCCACCAAG	GGCGGAGGTT	CATCGAAAAG	AGTCAAGGAT	AATTCAGGCA	TACTACCCGC	7 2 0 0
GACCGGGGCC	CAGAGGGCTA	GAATAAGCAT	TACAAGGTTC	AT		7 2 4 2

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18994 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGCTTTATT	ATGAGACATC	ATACACATAG	TATAGGCGAG	GTGATGGGGC	GGGGAAAGAG	6 0
TTGGAACCGA	AAGACAAAAA	AAAAAGCCTA	GTCGTA CT CG	GGATCTCTGA	GCGAGACGGG	1 2 0
TTGCATGGCA	ACTTTCATTA	GTTTGGGAAT	CTGCCAGCTG	GTGCTGTTCG	AAGGTTCTTC	1 8 0
CATTTCCGAG	GCGGTCAGTT	CATCGTACAC	CGAAACGTAG	TACCTGATGG	GGTCCTCCTC	2 4 0
ATTGTCCGAG	AGGTGAGATT	CGATGGTCAA	AGGCGAGCCT	CTCCATAAT	TGGGATTCAC	3 0 0
GAACGACGTG	TCCAAGTTGC	CATCCTTTCT	GAAATAGATG	ACGTTCTCAG	GATCATGTTT	3 6 0
CATGCGCTCG	CGGGCCGCGG	ACGCCTCCTC	CTCCTCGTCC	CAGTCCCGAG	TTTCCAACCG	4 2 0
CTGATAAGGG	CTCGAGGAAC	AAAATCCGGC	GGGGATCTGA	GAACCTCGTC	GGGAACCGCT	4 8 0
GCCAAACGGG	CTGCTGCCGC	CACTGTCTGC	CGTGTCTGTC	AACAGGTTGA	CGGCCTCTTC	5 4 0
GTCGGCGAAA	CGAAAGCGGC	CCGGGTGCTT	GCAACACGAG	GAGTAAACTA	CCGCGATCAG	6 0 0
TACCGCTATG	AAGCTGAAAA	TGGAGGTGCC	TGTCACGATG	TAGAAGAGGA	TAGCCAGCAC	6 6 0
TTTCATGATT	TCGTCAATTGC	GCGCGTCGTG	AACGGAAGAT	TCGCGGGCAG	TGGTCATGTT	7 2 0
GGTTTCCGGT	GTAGGTTTCGC	TACTCGTGGT	GCTCTCGACG	GTATTTCTGC	TGCTGGTGCT	7 8 0

-continued

AGTAGGGACG	TTTGTGCTGC	TGGTCATATT	TGTAGCGTCG	CTGAAGTCGA	TGTGAAGCAG	840
CAACCCGAAC	GCGACCAGGA	CCAGGAATGT	TGCGCGAAGG	AGACCCCGCG	GGGCCGGCAT	900
TCTTGAGACG	TGGCGACGTG	GATTTCTTGT	TATGTCCGCG	AACGACGTGT	AACGAGGACG	960
TGGTTTCCGC	AAGCCTCTAC	CGACGCCGCG	ACACCAGGTA	GGTTATCAAA	ACGCGAGCCC	1020
ATATCGCCGC	CATCATTGTA	ATCAGCAATG	TGTTGAGGTA	CTGCACGATG	AATCTGTCTA	1080
GTGACACCAG	CCAACCCTCT	GCTTTTGC GG	GCAAGCGCGC	TTTCGGTGAC	AGGGTGTATC	1140
GTACGTAGCC	GCGGGTCAGG	CGCGCGTTGT	AGCGGTACAC	GCAGAAATCT	ATCCACAGGC	1200
CAACGCCCGG	CTGTAGCTTC	GGATGGTGGA	TAATAGCGCG	GTGACGTACG	CCGCGTGGCT	1260
TTAGAATCTC	CACCTGTAAG	GCCATCTCCT	CCAGGTAGTG	GGTCTGACTG	CGACGCAGCG	1320
TCCAGTTCAT	GTAAAAGTCG	GTCTCGCCGT	GTCCGGCCAC	GAAGAGGCTG	CTTACTAATC	1380
CAGTCTACAT	TGTGCCATTT	CTCAGTCTGA	TTGCATTTTT	TAGAGTTATG	TTGCCACCAA	1440
CGTATCCTGT	TACGTTGATT	ACCTCGTAAC	TGCGGTTCGCA	TCTTTTATGG	ACTGTATAAT	1500
TGAAACCATC	ACAAGTAACG	GTCGTGGTGG	TGTTGGTACA	TGTGGTAGTC	TCAACGTTTG	1560
TATTTGTCGT	TGTGGATATC	TGTGTGGTTG	TTTTCGACGG	TTTTGTAGAA	ACGGTGGTTG	1620
CTGGTGCAGT	TGCAGTAGAG	CAATTTATAG	ATTCTGAAGT	GCTTTTATTG	CTGATAACTG	1680
TTGTACTGTA	TGTTGATGTG	GCTGTCTCAG	TACTAGTGGA	ATAGTTAACG	GTAGTACTAC	1740
AGGGACATTG	ACAGGAACAG	GTTTGATTGC	AGCTTTCTGA	TAACGCGGAT	ATTAGTATCG	1800
TCCACATAAC	CGTAAATCGC	CAGTCCATTG	CAATATTAGT	TCTCGCTCAA	TGGGCATTAA	1860
TATTCCTTTG	AACGCTGAGC	CTTACAGAAT	GTTTTAGTTT	ATTGTTTCAGC	TTCATAAGAT	1920
GTCTGCCCGG	AAACGTAGCT	CAATCTTCAT	GTTCTGTGTG	ATATCGAACA	ATGAATTCTG	1980
ATTCACTGAC	GGTGTCTTGC	AACATATGGT	ACTTTTTGTT	AAAGGCTCGT	CGTGCAAAAA	2040
ACAGAACTAT	GCAGGCCATA	CAAACCAACA	CGATGGTCCA	TACGGTGC GG	CTTCTTTGCG	2100
AGCTATGATA	GAGATTACGT	TTGTGGTGAT	GTGTATTGTT	AGTATGTTGA	CTTCCTTTCT	2160
CTCTATCTTC	ATTTTCGATA	TCGGTGTTGT	ATCTAGGGCA	AACGAAAGTA	GCATTAATAG	2220
CTTCAGTATG	ATTTTTTGGT	GTTACTAATA	GGTAGAAATT	TTCATCTTCG	TGATGTCCTG	2280
TGAAGTAATT	TTCTTTAAAA	CAACGTCTGC	TGTACCTGCC	GGAATTGGTG	ATATTTAGAT	2340
CGTACAGATG	TAGTTCTGTG	TTGTTGCACG	AACGACATAA	GTCATGGTAC	AGTGAAGGTC	2400
TTTCATGTTG	AGAATGACAT	ACAGTATGAG	AGGTAAGGAT	GCGTAACCAA	TACCAAGATT	2460
GGGTATGTGC	GTTCTTACGA	TGACCTAGAT	GATGTCCGTG	TGTGGATCGA	TTGTAATGTC	2520
GTATCCAGGC	GACTGAAAGA	CAATCCCACG	TAGAATTACC	TTTTATGGTG	ACATTACTCC	2580
CTTCTATTCC	TGTTGTATTA	GTTTCTTTGA	AACGTATGAT	TGTTGTCTCT	GTGTGACAAG	2640
CGTTGGAAGA	GTTAGTACGG	TTGTACGTGG	TGTACGTTGT	GGCGCTGCAA	TTTGTAAGCC	2700
ATGGCGTGCT	TATAAGTGCA	GTATTAGTGG	ATACGTTGTG	CGAAGTTTCA	TCTGACGTGA	2760
TAGTTACGGT	GATTGTTGTG	TTATAAGATG	ATGTAGCGTT	TGCTGTTACG	TTGGTAAAAG	2820
ATAATATAGT	GTTGGTATTT	GTTGAAATCA	ATTCTGTAGT	GGCAGCCGTA	TTGGATATAT	2880
TAGCATATGA	TGTATTGAAT	GTAGAATATA	CGGTTGTGAA	AGTACTCAAG	TCGGAGGTAA	2940
CGTTGGTGAT	GTTGCCAATG	GTTGACGCTT	GTGAGGTGAC	AGATGTGTGT	GGCGTCGTTG	3000
ATGTGTTGTT	ATTCGGAGTA	GAAAATACGC	TGGTCACAAA	GGTGGTAGAA	GCAGTGTGG	3060
GTGATGTGAT	GGATGCAGTA	TTGGTAGTAG	TACTGTTGCA	TGTAACTCTA	TGCAGAATAT	3120
AGAATATTAT	GATTGTATAC	GCCGTATGCC	TGTACGTGAG	ATGGTGAGGT	CTTCGGCAGG	3180

-continued

CGACACGCAT	CTTTTACTGT	AAATCCCCGT	CCACCGTCAA	CAACAAAGGT	TCCGTATCTA	3 2 4 0
GGTCCGTCCG	CAGATGTTCA	GCGTCCTGTT	CCCCGATTCG	TTGCGATCGC	AGGAAGCAGA	3 3 0 0
TGACCAGCGC	GCCAACAAAG	ATCATCATT	CCGAAACCCA	GGCGCAATGG	AGTGAGAGGC	3 3 6 0
CGGACCACTG	GCGTTTTAAA	TCCGAGATAA	TTGCCCGGTC	TGCCTCTTGG	GAATCCGTAA	3 4 2 0
CCACAACCTCT	CCCTGGTCCC	GGATAAAAAGC	ATCGACGCGT	TTCCAAGGCT	CGGCAGAAGC	3 4 8 0
TACGTGGGTG	GATGATGAGG	TAGAAAGCCT	CGACATCGCC	GGTATACTGA	TCCTGCAGGA	3 5 4 0
GGTAGACTCC	CGTATCTTTA	ACCGTGAGAT	TGTACAGCGT	CAGATTTTGG	CGCGTGCACG	3 6 0 0
CGAACGCCGC	ACCGCCCTGA	CGCGTGGTTT	CTTTATAGGC	GTCTGTAATG	ATACAAAGTG	3 6 6 0
GCGGCATACG	ACGCATGTAT	CTGCTGTAGA	TATCATAACG	CTGCCAGACT	ACGCTGTGAT	3 7 2 0
GGCTAGTGTT	AAGCCTGGTA	ACCAGCGTGC	GTGTACGGTC	CTCGCAGGTG	GCACGGTAGT	3 7 8 0
TGGCGAGCTT	TAGGGGTTTT	TTGGTTGGTT	CGACGGCGTT	CGATGAACTT	CCCTGAGTTG	3 8 4 0
TGAACAAAAA	CAGCGACGTG	ACTATGACAA	GCGTGAGGGG	GGTGCTGTAG	GTCTGCATGG	3 9 0 0
TGCAAAACAC	GTTCTCGCCT	TCCTTATCAG	ACGTTGTCGT	CCTCGTCCTC	TTCGTCTCT	3 9 6 0
GTGCCCGTCG	GTTTCGATCAA	CGGGGAGTTA	TCTTTCTGTC	TGGAGGGTCG	GTATGGAATC	4 0 2 0
CGTTCGTAGA	TGTTCTGCTT	TTTAGCCGCG	TGTTGTTCCA	GCTTTTTGCG	TGTCAGGCTC	4 0 8 0
CGATAGGCCA	GACATTGATC	TACCTCGGTG	CCCGTGTTGT	TTTTCTCCTC	CTCGCGCGCG	4 1 4 0
TAAATTACAA	AGAAGACCAC	CAGCAGGACT	ATCAGCGTAG	CCACGAACGA	GCCCGCGCCC	4 2 0 0
CAGGCCGAGT	ATGCGCCTAG	CATGGTAATG	GGTTCTGTGA	TCCGGCATT	GCACATCGCG	4 2 6 0
TGGCACTTGC	TGCCATTGCC	GGTATTAGAT	GATGTGTTAT	TCGGACTGCA	CTTGCACGTC	4 3 2 0
AAATGGGTAT	TTTCTGATTT	CACGAGACAG	TTGGTGGCGA	CTTTGGTTTC	GGCGCAGACG	4 3 8 0
GCCACATAGC	TTACCAAGCT	GAGTGCCAGA	AAGCACACCG	CGTGCATTAC	ACGCGGATAC	4 4 4 0
ATATTAAAAC	ACCGTGTTCC	ACAAGCACCG	CACACGTCAA	TCCTCCCCGC	ACGGTCTTCA	4 5 0 0
GCCCGCCCAT	GACATGATCT	CCCTCACGTT	ACCCTTCAAC	ACCCTGTAGT	ACTCTGTCTC	4 5 6 0
GGCTTCCGGT	CCCCATGTCC	TAATTATAAC	AAAACACCGT	GACACTGTCC	ATCTCCCTGT	4 6 2 0
CTTTTTGCGC	CGCCGGTCCC	CCCCAAATCA	TGTCTCTAGA	TGCCGCCGGC	CACCAACCGG	4 6 8 0
AGGCACGGCG	GCTATTGGAT	TCGGCATTGG	TGCGCCGCGT	CTTGGCCTGC	ATGATCATCG	4 7 4 0
TCATCATGAT	TGCCATTAGC	ATCTGGATCC	TGACCTACGT	GCTGTTTCTC	TAATAAGAAC	4 8 0 0
CCCGGCCCT	GACGGTAATT	TTCCTTTCTT	CTCCGTTTCT	CCTCAGCTGC	CGTACGTGAT	4 8 6 0
GCCTCACGGC	CATCTCCGAC	AGGCCCTCTC	CCCGACCTCC	TGGACATGTG	AGGGCTTGTT	4 9 2 0
GCTCCTCCTG	GGATTGCTGG	TGCTCTTCTT	TCACCACCAC	AACCAGTCGG	CCGTGGAGAG	4 9 8 0
GCGTCGCCGC	GTCTCGTTTCG	TCGAGGCCGA	TCGACTGCCG	CATGAGAGCG	GGTGGTATTC	5 0 4 0
TTCCGATGAC	GACGGAGACC	GGGACGGTGA	TGAGGAAACT	GGAGAGAGCC	ACAACAGAAA	5 1 0 0
CAGCGTGGGA	CTGTCCGCTG	TTTTTAGCTG	ACTGGCGTGC	GACCTGTAAA	CCGTTACTCG	5 1 6 0
GGTCTCAAGA	TGGTTTGGAA	GTTGTGACTC	ATCTTCTGT	GGGTGATACC	CAACCGGACG	5 2 2 0
CGAGTGTTCC	ATAAAAGCCG	GGCGCTCCGG	CGAGACCATG	CCATCCTCGC	CTTCGGACGC	5 2 8 0
CCCGCTCCTC	TTCTCTCTCC	TCTCCTCCCC	GCTGCCGCGG	CCATTGCCGC	CGCCGCCCAT	5 3 4 0
ACCATCGGCA	TGTCGGCCGA	CAAATCGCAG	CTGTCTTCGC	CGCCGCAGCT	GTAGCAGTTA	5 4 0 0
ACGTCGCCGG	CCTCCAGGAG	GAGATGGCGC	TGTGCGTCGT	CTCTTCGTCC	CGTCTCCCTC	5 4 6 0
TGTGGTCGTG	GGTGGTGCGA	GAGTACACGA	TGGGTGGCTC	TCGTCTCGGG	GGACCACAGG	5 5 2 0
GGGAGGGGGG	TAATTTATTA	TTCGTATTAC	TGTAATTTTG	TATCGCTTAA	TTTGTTTAGA	5 5 8 0

-continued

GCCGCACGCT	TGACAACGCC	TTGTATAGCC	TTATTTATCC	CGATGACTTT	TTTCTCCGTA	5 6 4 0
CAAGAAATGG	ACGTCACTTG	AGCAGACACA	GTTTCATCGA	CCACGACAGT	CTCATGATCT	5 7 0 0
GACTACCTCT	GACCCGCCAA	CGAGAAAACC	GAAAAGTAAA	AGATGACCGC	GCCCTCGGAG	5 7 6 0
TCCTTTTTTC	CTTTTCAATC	ATGAAAGCAA	GAGGCAGCCG	AGAGAATGCC	AGTAAGAGAC	5 8 2 0
GACCATCGCA	GACACAGTAC	GATACTCATC	TTAGAACGAA	CCAGCGAATA	ACCATCACAC	5 8 8 0
GTACAGCAGA	ATCTCATGAA	CTAGTCAACC	AACGTCATAA	AATCTTCACA	CAATCGTTTT	5 9 4 0
TGCGAACTTT	TAGGAACCAG	CAAGTCAACA	AAAGACTAAC	AAAGAAAAAC	CATCTTGGAA	6 0 0 0
TTAAAAAAG	TAGCATCGTT	ACCTTATGAA	CCAGCAGCAT	TCAGTATATA	CACCAGATAT	6 0 6 0
AATATATTTA	TTAATGTATC	CTCTCTTTCT	CCTGATGTAA	TTTTGTTTTT	GTAAATTCAA	6 1 2 0
TTGTTGAAAG	TCTCTCCCTG	GGGGAATTGC	ATATCTTATT	GATGAAGAAG	AAATCCCTGC	6 1 8 0
CATATGTGTT	GTCAAACAT	CATTATTTCT	CTATATGGGT	ATTTTTTTTC	TAAGAAGCAA	6 2 4 0
AAGACTAGCA	GCAGCCAAAA	TAAACCTGAT	GAAATCTTTA	ACTGAACTCC	CAGTGGTCTG	6 3 0 0
TGTGTATATT	TCTGTTGGTG	GTCGGTTGTC	TGAACCCGGG	TGGGTTGTTC	GGAAACGGCG	6 3 6 0
GGACGGGGAA	ACGGATGGAA	ACAGCGTCGC	TATATACGTG	ACTTTTGATC	TAAACGGACG	6 4 2 0
TCGCTAGGCT	GACAGTTTAC	GAATTGCTAA	ACAAGATAGG	AACAAAACAA	GCGGGGCTTT	6 4 8 0
GCCTGGTAGG	ATTTCCGTG	GAAACAATAA	CCGGATGTGA	TTGTGGCTGG	TACATAAGCT	6 5 4 0
GGTTCTGGCT	GCAAGCGCTT	TTCACTGCAT	TAGGTTTGGC	GTTTGCTTTT	GCCTGGGAAC	6 6 0 0
GCTATGGCTA	TAACGGGAAA	GAACCGGTTT	GGCAACATTC	CATTGTGGGG	GGGGGGTACT	6 6 6 0
TATAGCGTGC	CTAGCTATGA	CGTTGATATA	TGTGGATGCG	GATAATACTC	GTAATGAGCT	6 7 2 0
AAAAGCGACG	ACTGGTAGTA	ATTTTACCAT	TACGCATAGG	AAAGATCCGT	TGACAACATA	6 7 8 0
GTGGAAAACC	GTTTTTGGTA	ACAATGGTGA	TCAGTGGTTG	TGCAACGTTA	CGGGTATAGG	6 8 4 0
TAATGCTACT	GTGAATAGTA	ACGCAACTAT	TTGTGTGTCG	AGCTGTGGTC	ATAATACGTT	6 9 0 0
GGATTTATGT	AATTTAAAGT	CGGGAGATTC	TGGCTTCTTC	GATCTGTCTC	GTTGGTTCCG	6 9 6 0
TGAAAACATG	GATGAATACA	GTGGTGATGT	GTGGCACTTG	GAAGTCAGCT	AAATGTTGTA	7 0 2 0
TCGCTTAGTG	AATTGGTGTT	CTTACAGTTT	TCATGTAATA	AACTACGTGT	AATTCGTTAA	7 0 8 0
ATTTGTGTGT	TTTTTTGTTA	GTATTCTGCG	TAACGGTGGA	ATAAAATTGC	GTTGACCTAG	7 1 4 0
TTAGATTTCC	TGTGTAGAAC	AATGACCGGA	CGTGCTTGGA	CTGGTACATA	CGCAGGGGCT	7 2 0 0
GGACGTGGTT	ACCGGTCACT	GGACTCGGTT	TCGCTGTAGC	TGTGGTTCAA	CCTGAACATG	7 2 6 0
GCTCCCAGAG	CTGCTAGGAA	CCGGTCCAGT	CACATTTTTT	GGTGGGTGGG	GGGTAATAAA	7 3 2 0
AAAGTGTTTA	ATATTTGGGT	TTAATGATAA	AATCCAGGTT	ATGGATATGA	GGAAACTGAA	7 3 8 0
TACCTCGCAG	GGTCGAAATC	TTACCACAGT	TGATGATAGA	AGACGGTTTT	CCATCGGGTG	7 4 4 0
GGAAACATGG	GATAACGGTG	GTGACTAATA	ATGGTACAAC	GGTCGTCAAT	ACAACAGCCT	7 5 0 0
GTGTTTCAAG	TTGTTTCGCAT	ACGTCGCTTG	TGCTTTGCAA	TATGACGCAG	CAGACTGATT	7 5 6 0
CGTTGTACGG	AGTGGGTCAT	CGGTTGAATG	ACGAAGAAGA	TGGTGAACGT	TGGAGAGTTT	7 6 2 0
CGGTTTCTTA	ATAATCCCAT	ACGACATGTG	TTCATTTATA	TCTGAATTTT	AGGATGATGA	7 6 8 0
CTATAGTATA	ACTCTGGGGA	ACAAATATCA	TACGTTAATC	ACTTTAAGTT	ACGCCGTTAG	7 7 4 0
GAAAAGAAAA	TCAGTCCGAA	TGAAGCATAG	TCAGCCGAAT	GATACAGCAA	TAGCTTGTTT	7 8 0 0
ACAACGTGTT	CTTTTTTACA	TTATGAACGT	GCCTTGCTTT	TTATACACAC	ATGGAGACAG	7 8 6 0
AGGTCCCTCA	GCCCTTGTC	CGACAACCTC	CTTTTTCTAA	ACCGTATGTG	CTCCAAACCG	7 9 2 0
TATCTCCTCA	TCGTCACGTG	AAATACCATG	GGACCCCTTT	TCGTCACACA	CGTCTTTCCG	7 9 8 0

-continued

CTTACCCAAC	GCGTCAGCCC	GCGCTCGGCA	GAGCTACCAT	ATAAAAACGC	AGGGGTTTAG	8040
CAGCTTCCCC	AGATCGCTGC	TGCCCCGGCG	TTCTCCAGAA	GCCCCGGCGG	GCGAATCGGC	8100
CGGCTGGTCG	GTCGGCGCTC	GGACGGATGG	GGAGAACGGC	GGTGACTION	CCGCCCGTGG	8160
CCGGGAGAAAG	ACGGAGGAGC	CGAGATGACA	ACAGCAGTCG	TGGAAGGGTC	GCCAAGCCCC	8220
GGTCCTTCTC	TTCTGTCTGG	TCGAATCTTG	TTTTCTTTTT	TCAACCGCTC	TTTTTGTAC	8280
CTTTTTATGT	GAGTTTCTCT	TCCGCGTCTC	CCGGCCGTAC	CATCCACCCA	TGCAGCATGC	8340
ACGCGTGTAT	GTATGCATCG	CCTCTCCTCC	GTCCCGACTA	CCATCAGCAG	TACCACTGCC	8400
GCCACCCCCA	GCGCCACCAC	CGCTGCCGTC	GCCACCGCGT	TATCCGTTCC	TCGTAGGCTG	8460
GTCCTGGGGA	ACGGGTCGGC	GGCCGGTCGG	CTTCTGTTTT	ATTATTTTTT	TTTATTTTTT	8520
ATCTTCTCCT	TTCCTTAATC	TCGGATTATC	ATTTCCCTCT	CCTACCTACC	ACGAATCGCA	8580
GATGATAAAC	AAGAGGGTAA	AAAGAAAAAA	GCTACAGACA	TTTGGGTACC	TCAGCTTTCC	8640
GATAACTCGA	AGAATTCAAA	GTCGACGATT	CCCAACAAGA	GAAAACAGAA	CAAAAACAAG	8700
GTCATTTTTA	TTTATCCTCA	TCGTCAACAA	CAACTACCGA	CAACAACGAA	ACACCACCAA	8760
GAATGTCAAT	CCGCAAGGGT	GTTCCTGCCC	CCTCGACGCG	CCTGTCGCGA	TCCTCATGGC	8820
GAGGACCGCG	ATCTCCGTAT	AGGTAGATGA	AATTATCCCG	TGTCCGGTCC	TGATTCCCCG	8880
CATGCCCTGC	ACATCCTGAC	GCGTCGGTCA	GCAGCCAAAC	AATCATAGGA	AATGAACCAG	8940
AAGAACAAAA	AGATCATCTC	TCTCGGTGTA	TAGCAACACC	AACAACAACC	GCATCGCAAC	9000
ATCTTCATCC	GCAAGACGGA	AAGAAAACAA	CAATAATGAG	AATGAAATCA	CCACAACCAA	9060
GCCAGATTTT	ACGTCCATGA	GTTTTTATTA	TATTATTATC	AAAACGAAAA	ACAGAAAAAC	9120
TGTCATAGAT	AAATATAAAA	AAAAATAGAA	ACCACAAACG	ACTACTAGTA	CTCCAATCTT	9180
AGATGTATAT	GCTCCTAGAT	AAGATTTAGT	ATTACCATAA	TCATCGAAGA	ATGAAAGACG	9240
ACGATGATTC	CTTACCGCTC	CTGCCACCCG	GTCTGTATGT	AGAGAGAGAA	GAGAGAAAAC	9300
GGTGAATCCA	AGATCCCCGG	GTCGGCGTCG	GCATGCCGCT	GATCGCAGTG	GCCCCACCTC	9360
GGCATGCCGG	CGCCGGGCGA	GGAATTGCTC	ATGAAAAAAA	GTATCTTTCT	GTAAAAAAG	9420
AAAACAATAC	ATGATTAACC	GAAAAGAAAC	CAACAAAAAAG	AACCCGAGAT	CAGTCGATTT	9480
CGATCACTAC	GATAAACACA	TGGAAGATTT	CTTGAAAAAA	GAAAAGAGAA	AGAGACCACC	9540
TTCCCGGCGG	CGGACACGCT	CCTCTCCGTC	GCCGTTCTGC	ACCATGATTC	GATCAATAAC	9600
AACATCATCA	TCGGAGACCA	TCTTTTAATC	AATCAGCGTT	GCAGTAGTCG	ACTCCCTGGA	9660
CACGAAGGAG	TCATCCATTT	TTATCCTCGC	ACTTCTTCGC	TCTCAAAGCC	GCCTTTAAAG	9720
TTGAAATGAA	AGGATGGAAA	CATGGAATAC	AGTTTTAATT	GCACGTATCA	CCATTTTACT	9780
ACAAAAAGAA	AAAAAAACAA	CTTACACATA	GTATTACCTT	AGGTTTACGG	ATAAGTAGAG	9840
TGTAGGCGTT	TTTGAAACAG	TTCAGCCAAT	GCAATCTTGT	CTCGGCATAA	TCACTCTTTC	9900
TGCATATAAT	AGTAGTAGTA	GATTTATTCA	CATCAACACA	GCGAAAAACT	CCAGCATCAA	9960
AGTACACCTA	GAGACAGCCC	TTAAAATATA	GTTTGCAGCT	TTTAGATGTA	CTTACACCAA	10020
AGAAGATTAC	CGTCCTTACG	AGAAAACAGA	TACTCGGATA	TAGGAATCAA	GACAGCTCTG	10080
CACTGAAAAC	ACACTCTCCT	GTCACGACAC	CGCGCCACAC	CAGAGGCGTA	CGCGTGACTT	10140
CATCGCAACG	ATCCATCGTG	ATGTCCCTCG	CAGAACCTAA	AAAGACCAA	AAAAAATCTT	10200
GGACCACAGT	TGTCGATACT	TGAAGACAAT	ATTCTCGTGA	GAACCTTGAG	ATTCGCACTT	10260
GAAACCTCTT	AGGATCCACA	AAAACAACAA	CCTCTGTATG	GAAAATGCGC	TATTTTATCT	10320
CAGCTTTTCT	CCCAAACCTC	GGTTTCTTCC	TATTCTTATG	TTTTCCCTAG	TATATTTGCC	10380

-continued

TCCTTATAAG	AAAAGAAGCA	CAAGCTCGGT	CGCACGGATT	ATTCCTTCTG	CTAATCTATT	10440
ATTTTGTTC	TTTTTTTTTT	CTTTGCCTTC	ACCCTCTTCA	CTCCCTGTAG	CAACACAGAG	10500
TAGTAGACAC	AATAAATGAG	AAGTTTGCAT	GCATTTGTCTG	TGTCCGTGGT	TTGTTATGGC	10560
GTGTGGAGTG	CTCGGGATGG	GTGGACGTGG	GGACGGATTTC	TTGAGGCTAC	AAAGATACGC	10620
GGAGACGTCTG	TGGCGAGGGG	ATGGGTTTAT	TGGATATCGG	TGAAGCAGCG	TGGCGGCGAA	10680
AGACGCGATC	CCTGGGCTGG	TAGATCCCCC	TACCCCGTCT	ACCAGGGACG	TTTATCCTTT	10740
GGACACGTAA	ATGTCTCGGC	CGGCATCCAC	GCGCCACGTT	CACCGCGTTG	TGCCCAGCGC	10800
CATGTGCGGG	TCGTTTCGGC	GTGAAGTTGG	ACGGCGTAGT	TTCGGGGATT	GTGAACCGTG	10860
GCTGAGGGTG	TAGATGGGAC	AGGAAAAAGC	GTGTGATCTG	ACCGAGGCGA	AGCATGTGGG	10920
TGGTGCGATG	CGGTGGATGT	GGCGGGGTGC	GGCGGTTTCC	GACGTGGAGA	TGTGGAGATG	10980
GGGGTGATCC	GGATGCGTGG	CAAGAGGCCT	CGAGCTTGGG	CTTCTCCCGC	GGATGGACGT	11040
TCTAACTGTA	CACGGCGGCC	GTGGCCTCCG	AGTAAAAAAA	CCAGGTGCTG	ACGCCAGACA	11100
GAGACGCCGT	CCTCGGAATC	GTGTGCGCGA	AAGCCTGTGC	CGCGGCAGCG	TACGACGTTT	11160
CAGTCAGCGA	GGCCGTCGCG	TTGGCGCGCC	AACAGTAAGG	TGACGACAGG	TTGGCGGCC	11220
ATGGTTCCGA	AGCGTCCCCA	CATGCACCAG	CAGTCGGCGT	CAAAGTCGCT	TGCGCTGTCTG	11280
GCCCAGTCGC	CACCGCCGCG	GCGGATTTCC	GCGCGGGGGA	CGGGGTAGCC	GAGTGCTGCG	11340
CCCTCGCCAA	TGTTGTGAAG	TGGATGCGTG	AGTTGATGTT	GATTCTCTGT	GGGAAAATGA	11400
GCGCTGTCCT	GTGGGTTGGT	GTTGGGGTAT	GCGAGTAGTA	GGGGTTGTGT	TTGATCGTAG	11460
AGGTGTTGGC	GGGCCTGTGC	GCAAGCAGCG	TAGTCTGCGG	CGTCGAGCTC	CATCTGTGTG	11520
CGGTGTTCTT	CGTCGGCGTG	TTTGTCCGAG	GTTTGGACAT	GCGGTTGTGT	GTTGCTGTGG	11580
TGTAAGGGTA	ACGTGTGTTG	GGCGTCTGGG	TGAAGCGGCG	TGGTGTGGGT	GCTGTTTGTG	11640
TCTGTGGCTG	GCATGATTGT	GCGGCATGTG	TGTGTTGTAG	TGGGTGGAGG	TTAAATAGGT	11700
GAGGTGGGTT	CCCTGGTCCG	CGCCGCAAAC	TGTCCCCGTC	CCCAACGTAA	CCTCCCCTAC	11760
GCGGC GCGAA	CAGCCCCGGC	CCCAGCGCAA	CCCCCGTCCC	CGGCCCCAAC	ACCGTCCC GC	11820
ACACCCCCCG	TCTCCGCAAC	ACCCCGGCAT	CGCCGGCGGC	CAGAACGCTC	GAAAACCCCC	11880
GACAAGCGCA	GCGCCGAAAC	GACACAGGCA	AGGACCGTGG	AACGCACCGG	CAGCGCGCCG	11940
AAACACCGTC	CCGAAGCCCC	GTGCCGACAA	CAAATACCGT	GGGACGACAC	GCACCGGCAG	12000
TGCGCAGGCA	GCGTCCGACA	CAACACGCTT	ACGGCCCTCA	ACACTCCCTC	GAGGACCCAC	12060
CACGCGGCC	CGCACCGGCG	GTGTTTTGGG	TGTGTCGGGG	CGCGGCCGGG	TGGGTGTGTG	12120
CCGGGTGTGT	CGCGGGCGTG	TGTTGGGTGT	GTCGGGGGTG	TGTTGGCAGG	GTGTGTCAGG	12180
GTGTGTCGCG	GGCGTGTGCC	GGGTGTGTCTG	TGCCGGGTGT	GTCGCGGGCG	TGTGGCGGGT	12240
GTGCCGGCGG	GGTGTGGTGG	CGGGGTGTGT	CGGCGGTGTG	CGCGGCCTCG	GGGTGTGCGG	12300
CTTCGCAGGA	ACGAGTGTGT	GGCCTCGCGG	CCGTTATTTT	CCCCGCGGTC	CCCAGGGCCG	12360
TCGTCCCTCG	CCCCCGGGCG	TTGCTTTTTCG	TGTGTCCCCA	GGGACCCATG	CTGCCGTCCC	12420
CCGGGAACTT	CCTCTTTTTC	CCGGGGGAATC	ACACAGACAC	AGACACGCGT	CTTCTTTTTCG	12480
CCGTGCGCGC	CGCACGTCTGC	TTTTATTCGC	CGTCGCCGTC	CTCCGCACCA	CACGCAACTA	12540
GTCGCCGTCC	ACACACGCAA	CTCCAAGTTT	CACCCCCCG	CTAAAAACAC	CCCCCGCCC	12600
CTCGAGGACC	CACCACGCGG	CCCGGAATGG	ATGTCGGGCG	TCCACCTAGA	TGGGTGCGCG	12660
CCCGGGAGGC	GGCTGTGCGC	TCCAGTGGTA	CGCGCCTGCC	GCGCGTCTTC	CTTCGGGTAG	12720
CTGCCTTTCC	CAGTCCACGG	CCTTCCAGAC	TGCGTGGCGC	CAAGGCGGCG	CCAGCACGCG	12780

-continued

CCGTGCACGT	CGCTGCCTAT	AAAAGCCAGC	TGCGTGTCGC	CCGCGGCACA	CGGGCGACGA	1 2 8 4 0
AGGCGTCCGC	GTGTCTAAAC	CGCGTGCTCG	CTGACGCGGG	TTTGCTTCCT	ATATAGTGGA	1 2 9 0 0
CGTCGGAGGT	GTCCGGCGCC	CATGGCCCAG	CGCAACGGCA	TGTCGCCGCG	CCCCCGCCC	1 2 9 6 0
CTTGGTCGCG	GCCGCGGGGC	CGGAGGGCCT	TCGGGGGTTG	GTTCTCTCC	TCCTTCTTCT	1 3 0 2 0
TGTGTGCCGA	TGGGAGCGCC	GTCAACAGCG	GGCACTGGTG	CGAGTGCTGC	GGCTACGACG	1 3 0 8 0
ACGCCGGGCC	ACGGCGTCCA	CCGGGTAGAA	CCCCGCGGGC	CGCCGGGCGC	CCCTCCGAGT	1 3 1 4 0
AGCGGCAACA	ATAGCAACTT	TTGGCACGGC	CCGGAGCGCC	TGTTGCTGTC	TCAGATTCCG	1 3 2 0 0
GTGGAGCGCC	AGGCGCTGAC	GGAGCTGGAA	TACCAGGCCA	TGGGCGCCGT	GTGGCGCGCG	1 3 2 6 0
GCGTTTTTGG	CCAACAGCAC	GGGCCGCGCC	ATGCGCAAGT	GGTCGCAGCG	CGACGCGGGC	1 3 3 2 0
ACGCTGCTGC	CGCTCGGACG	GCCGTACGGA	TTCTACGCGC	GGGTGACGCC	GCGCAGCCAG	1 3 3 8 0
ATGAACGGCG	TGGGCGCGAC	GGACCTGCGT	CAACTGTTCG	CGCGGGACGC	GTGGATCGTA	1 3 4 4 0
CTGGTGGCTA	CCGTGGTGCA	CGAGGTGGAC	CCCGCAGCCG	ACCCGACGGT	GGGCGACAAG	1 3 5 0 0
GCCGGCCATC	CCGAGGGTCT	GTGCGCGCAG	GACGGACTGT	ACCTGGCGCT	GGGCGCCGGG	1 3 5 6 0
TTCCGCGTGT	TCGTGTACGA	CCTGGCAAAC	AACACGCTGA	TCCTAGCGGC	GCGCGACGCG	1 3 6 2 0
GACGAGTGGT	TTCGGCACGG	CGCGGGCGAG	GTGGTGCGGC	TGTACCGCTG	CAACCGGCTG	1 3 6 8 0
GGCGTGGGCA	CCCCGCGCGC	GACGCTGCTG	CCTCAGCCGG	CGCTCCGACA	GACGTTGCTG	1 3 7 4 0
CGCGCCGAGG	AGGCGACGGC	GCTCGGACGG	GAGCTGCGCC	GGCGGTGGGC	CGGCACGACG	1 3 8 0 0
GTGGCGCTGC	AGACGCCGGG	CAGGCGACTG	CAGCCGATGG	TACTGCTGGG	CGCGTGGCAG	1 3 8 6 0
GAGCTGGCGC	AGTACGAGCC	GTTCGCGTCG	GCGCCGCACC	CCGCGTCGCT	GCTGACGGCC	1 3 9 2 0
GTGCGTCGGC	ACCTGAACCA	GCGTCTGTGC	TGCGGCTGGC	TGGCGCTGGG	CGCGGTGCTG	1 3 9 8 0
CCCCGCGCGT	GGCTGGGCTG	CGCGGCGGGG	CCGGCGACGG	GGACGGCGGC	GGGGACGACG	1 4 0 4 0
TCGCCGCCAG	CGGCGAGCGG	CACGGAGACG	GAGGCCGCCG	GCGGGGACGC	GCCGTGCGCG	1 4 1 0 0
ATAGCGGGAG	CCGTGGGGTC	CGCTGTACCT	GTGCCTCCGC	AGCCGTACGG	CGCCGCCGGC	1 4 1 6 0
GGGGGCGCGA	TTTGCGTGCC	TAACGCGGAC	GCGCACGCGG	TGGTCGGGGC	GGACGCGGCA	1 4 2 2 0
GCAGCAGCGG	CGCCGACGGT	GATGGTGGGT	TCGACAGCGA	TGGCGGGTCC	GGCGGCGTCG	1 4 2 8 0
GGGACCGTGC	CGCGCGCCAT	GCTGGTGGTG	CTGCTGGACG	AGCTGGGCGC	CGTGTTCCGG	1 4 3 4 0
TACTGCCCGC	TGGACGGGCA	CGTGTACCCG	CTGGCGGCGG	AGCTGTGCA	CTTTCTGCGC	1 4 4 0 0
GCGGGCGTGC	TGGGCGCGCT	GGCGCTGGGA	CGCGAGTCGG	CGCCCGCCGC	CGAGGCCGCG	1 4 4 6 0
CGGCGGCTGC	TGCCCGAGCT	GGACCGCGAG	CAGTGGGAGC	GGCCGCGCTG	GGACGCGCTG	1 4 5 2 0
CACCTGCACC	CGCGCGCCGC	GCTGTGGGCG	CGCGAGCCGC	ACGGGCAGTG	GGAGTTCATG	1 4 5 8 0
TTTCGCGAAC	AACGCGGTGA	CCCCATAAAT	GATCCCCTCG	CATTTCTGCT	TTCCGGACGCT	1 4 6 4 0
CGAACTCTCG	GTCTCGACCT	CACCACCGTC	ATGACAGAGC	GTCAAAGTCA	ATTGCCCGAA	1 4 7 0 0
AAGTATATCG	GTTTCTATCA	GATTAGGAAA	CCTCCTTGGC	TCATGGAACA	ACCTCCACCC	1 4 7 6 0
CCATCTCGCC	AAACCAAACC	GGACGCTGCA	ACGATGCCCC	CACCGCTCAG	TGCTCAGGCA	1 4 8 2 0
AGCGTCAGCT	ACGCGCTCCG	ATACGATGAC	GAGTCCTGGC	GCCCGCTCAG	CACAGTTGAC	1 4 8 8 0
GACCACAAAG	CCTGGTTGGA	TCTCGACGAA	TCACATTGGG	TCCTCGGGGA	CAGCCGACCC	1 4 9 4 0
GACGATATAA	AACAACGCAG	ACTGCTGAAG	GCCACTCAAC	GACGAGGCGC	CGAAATCGAC	1 5 0 0 0
AGACCCATGC	CTGTGCTGCC	TGAAGAATGT	TACGACCAAC	GCTTCACTAC	CGAAGGCCAC	1 5 0 6 0
CAGGTCATCC	CGTTGTGCGC	GTCCGAACCC	GAGGATGACG	ACGAAGATCC	TACCTACGAC	1 5 1 2 0
GAATTGCCGT	CGCGCCCACC	CCAGAAACAT	AAGCCGCCAG	ACAAACCTCC	GCGCTTATGC	1 5 1 8 0

-continued

AAAACGGGCC	CCGGCCCACC	TCCGCTGCCG	CCAAAGCAAC	GGCACGGTTC	CACCGACGGA	1 5 2 4 0
AAAGTTTCTG	CGCCCCGACA	GTCGGAGCAT	CATAAAAGAC	AGACCCGACC	GCCAAGGCCG	1 5 3 0 0
CCACCGCCCA	AATTCGGGGA	TAGAACCGCG	GCCCATCTCT	CGCAAAATAT	GCGGGACATG	1 5 3 6 0
TACCTCGATA	TGTGTACATC	TTCGGGCCAC	AGGCCACGGC	CGCCAGCACC	TCCGCGGCCG	1 5 4 2 0
AAAAAATGTC	AAACACACGC	CCCTCACCAC	GTTTCATCATT	GAAAGTCTCT	CCAGTCCATA	1 5 4 8 0
TGTTGTCAGG	ACGTGCTGTC	GTTCTCCGCT	TGCTGCGAAG	CCCGTTCTTC	CGAGTCGTGT	1 5 5 4 0
CGCTGCGTCC	AGCGTCGCGC	CCAAGATGGG	AATTTGGGTC	TTTTACGCG	TAGCCTCCTC	1 5 6 0 0
CACCACGGCT	GCTGATCGCC	GTCACTAAGG	ACCGACACGG	AGGATGACGA	GGAGCTTCTC	1 5 6 6 0
CCCGACTCCG	CGGTCCGCGA	CCGGCTACGT	AGCGCGTGTC	CCTGCCAGTC	TCCGCAGTTA	1 5 7 2 0
CACCACACGT	CGTGAGCAGC	GTGCACCTGC	TGCCGCCACT	GGGCCTCGGC	GTGCTCAGGC	1 5 7 8 0
CACCCGCCGG	AGCCCGGTCT	GAGCTCCGAC	GCAGGATGCG	CGTACTCAAC	GTGCGCCTTC	1 5 8 4 0
CAGTCCATAC	AGCAACACCA	TAGGTCGTGC	GAGTCGTGCG	CTACCCGCCG	CCAGGCCAGT	1 5 9 0 0
TCCCGCATGG	GAAGGCTGGA	CACGCCGACC	GAGAGGTCAC	CGAGCCCGGA	CGCCATCTCT	1 5 9 6 0
TCTTCCTCTC	CGTCGCTGTC	ATTAAGCAGC	CAGGTCACCT	CCTCCGCTCC	GCGTCCGCCG	1 6 0 2 0
GTCTCGACGG	ACCGCGCCGC	CGTCGGCAAC	ACGGAAAACA	GCACGCCAGC	CCGAGCCGCT	1 6 0 8 0
AAGGCCGCAT	GCCCCTGCCG	CCCAACTGAA	CACGCATACC	CCGCTCAACT	GCGTTTTGCC	1 6 1 4 0
ACCCCTGTCA	GTGCTCTCGC	TCGAGCACCA	CCCCGCATCT	CCCAACCTTT	TTCCAATAAA	1 6 2 0 0
CGAAACCGAC	ATGACACACG	TAATGGGTAC	TCGTGGCTAG	ATTTATTGAA	ATAAACCGCG	1 6 2 6 0
ATCCCGGGCG	TCTCAGCACA	CGAAAAACCG	CATCCACATC	ATAGACAAGT	TACAGTCCAC	1 6 3 2 0
AGTCACATAC	ACGATAAACA	ATACCAACAG	GGTAATGTTT	ATGGAGTAAA	ACACTATTGT	1 6 3 8 0
CCAGGCCACA	TGCGTGTATG	ACTTCCGCAC	CATCCCGTAC	TGCATGTTCC	ACATGTACGC	1 6 4 4 0
GCTAGACGTG	TAATCCACTC	GCAGTTCGGG	GACGCAACGC	AGCCAGATCA	CATCCCCTTG	1 6 5 0 0
CAGTACCAGA	CGCAGGGCTA	GCGTCTCGAA	GATCGGCATC	ACATCTAAGT	TCCGCACGTT	1 6 5 6 0
CCACTTTAAC	GACTCCCCGG	GAACGAACTC	CACGTGTCG	GCGTGTACGT	ACAGGTTCTC	1 6 6 2 0
TCCCACGCCG	CCATAATCGG	CCTTCGGATC	GAAGACGAAC	CGACTCATGT	TGCCCACGAT	1 6 6 8 0
GCTCCCCCGA	GCAAACAAC	TGCCGTTGTC	AATGTAGCAC	CGGTTGTCCT	CGATTTGAAA	1 6 7 4 0
CCAGGGATGC	TTGGCCGTGG	ACTTCCAGGG	CCGGAGCGCG	TCTTCCCCGG	CTTTAGTGAT	1 6 8 0 0
TCCATCGGGC	AGGCGGATCA	AGGGACCCAT	GGAGGTCCAA	AGACCCACCC	AGGCTTTCCA	1 6 8 6 0
GAGATTGTTT	ATGGTGAAAC	AGCGTGTGGA	CTGTACGCTC	TTTCCCAATT	TATATCCCAG	1 6 9 2 0
AGTAGTGACG	TGAGCCCAGC	CACCTCCCAG	ATTCCTGACG	TTTTGGTTGT	CTTTCCTGCC	1 6 9 8 0
AATTCCTCCC	GTAAACTTAT	GATTATCCTA	GCCCATTCCC	GATAAAAATA	CACGGAGACA	1 7 0 4 0
GTAGATAGAG	TTACGAATAA	ACCGGTTTAT	TTATTCAAGT	GTCTCAGGAG	ATTATTGAAC	1 7 1 0 0
GAGCGTGGAT	ACCACGCCGT	CGTCAGTTCA	TGGTGGCATT	GAGCAGCCAT	AGCACCAGAG	1 7 1 6 0
TCCCGGCGCC	CGGTATCAGA	CACGCTGACC	TACCGGGCGC	CTTCGAGTCC	GTACCCCGCG	1 7 2 2 0
GCCTGGGTGT	TAGAGTCCGT	ACCTTGCAGC	CCAGGTAGGT	TTCAGGTACC	AGCTGGTTCC	1 7 2 8 0
TACCTGTTAA	ATAAATCGCA	GACGGGCGCT	CACCCCTACG	GTCAGGAGCA	CAAGAACAAC	1 7 3 4 0
CAGAGAGAAC	AGATATACGA	GCAGGGTTCT	GAACAGCAGA	CCCCAATTGT	CGTCTCTCAT	1 7 4 0 0
GCTTCGCTGA	AGGTACCAGT	TGATGGTCTG	AGAGCTATAG	TCCATCCTCA	CCTGAGGAAC	1 7 4 6 0
ACACGCGGCA	TATTTCTTGG	GGTCTCCCCA	CCTCGTAGAC	AACGTGATGT	CCACCATATC	1 7 5 2 0
CACGGTGTGC	GTCACCGGGT	GCCCACCGAT	GTTCCACTCG	AAATAGGCTC	CGCGCTCATC	1 7 5 8 0

-continued

ATGGTGGTAC	TGCTCACCGG	ACACCTGCAG	TCTGTCCATG	TAAGATTGAG	AGACGATACC	17640
CACGTTTACA	AAGTGTTTCT	CGGTGAAGTT	GCCCGACATC	CTCCCCTTGA	AGTACAGCAT	17700
GCCCATATGG	AACCAGCATT	GGTTCTCCTC	CACTCGAAAG	TGGGCCGATC	TGATCTCCGA	17760
TACCACCACA	TCCAGGGGCC	GGGGCACCGA	GTCCGCGAGT	CTCAGGAACA	AGACGGCCAG	17820
GATCGCGAGC	ACCAACACCG	GCTTCATGGC	TCCGAAGGTC	CGCTGCTCGG	CTCCGCTCAC	17880
CGCTCCGGTC	TGGCTGCAGC	AGTGCTTCGC	TGAGAAGTAG	CGTGTGGACT	GAACGGTGTT	17940
TTTGAATATA	TAGCGTTTCT	TGGTGACGTT	GTTTCCCCTA	CGTAGTAGGC	AACTACGTGC	18000
CAAAAGAGGC	GTTACGGTAC	TTTCCGTA CT	GGGATTTCCA	AACCGGGACT	TTCCACACGG	18060
CGGTTTCAAC	ACCGGGACTT	TTACACACGGT	GATTTTCGGCA	CCGGGACTTT	CCGCACGGCG	18120
GTTTCGCCAC	CGCTGACGTT	CTCATCGCCG	CCCACGTCAA	CGGTGGCGAC	ACCGTACTTT	18180
CCCATGCGGT	TTATAAACGT	CAAGAGTCAC	GTCAGTCGCC	CACCCCCATT	ACACGGCGAT	18240
ATCCCGATAG	GGCATGAGGG	GACCCGGGTG	TCGCGACATG	TCGACGACAG	GTGCGGATTA	18300
GTGGTCGTGT	CGCGACATGG	ACGTGCAGGG	GGATGTCTGT	CGCGATAGAG	TTGATGTGAC	18360
AGCCCGCTAC	ACCTCTCTGT	CGCGACATGC	ATACACAACG	GGCCGGCTTG	TCGGCGATTG	18420
TCGCGACATA	TCGTTATCAG	TTAGCGACCG	GAGTTGTCTA	TCGCGACATA	TCGTCTGACTA	18480
TCGCGACAGA	AAAAATACCG	TTCGTAGAGA	ATGCCGTGTT	GAAGGAACGC	GCTTTTATTG	18540
AGACGATAAA	ACAGCATCAG	GAGCCACAAC	GTCGAATCCC	ACGTCCAGTC	GATTCGTATG	18600
TTATGCTGCA	CAGCAATGCT	AGAATAACAA	CCAGCAGGGT	AATCCCGCAA	CATAAATACA	18660
AAGTCACAGC	GAAGAATCCG	TGTCGTTCTA	TCAAGCGAAA	CGCGTTCCAA	ACGGCCCCGT	18720
CACAGACGCA	GTTATTCATA	AGCGTTAACA	ACCGGTGGCT	AGGATGAATA	TCCAAATCAC	18780
AGGGCAGTAG	CCGACGGACT	CGTTGACAGG	TCAGCCTACC	CTCAAGGTTC	CTATCGTTTCG	18840
GACGGGATTT	GTGCGTTTTA	GGCCTCTTTT	TCGCCGCCTG	CAAGCATTGG	TGCGCAAAGT	18900
CCTCACCCAG	CTGTTTCCAG	CTATCATCTG	CATCTGTGCA	GTCCCCTGTA	TCGTTGTAAC	18960
AAACGGGTCT	GTGCGACTTC	GTTCTCGGAA	CACA			18994

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5020 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (i i) MOLECULE TYPE: DNA (genomic)

- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGCTTGTG	TCGCGGAGAC	AGAGAGAGAA	GGGTTTTTCGG	GTCACGCGAA	GACCGCTCAC	60
CGGGGGTCGG	CAACGCACAC	ATCAACAGAA	AACCGAGACG	AATCAAGAGA	TCCATAGTGA	120
AGGAGTGATA	TCGACGTGCT	TACGAAACGG	CGATTATATA	TGTTCTCAAC	AATACCGCCC	180
TACGTTGTAT	GATGTAACGT	GTGACGTGAG	TCTGATCCAA	CACTGAACGC	TTTCGTCTGTG	240
TTTTTCATGC	AGCTTTTACA	GACCATGACA	AGCCTGACGA	GAGCGTTCAT	CGGGGCATGA	300
AGTACGCATT	ACACAAACTC	CATATATTTG	TTACGATAGA	ATACGGAACG	GAGGAGGCTT	360
TCGCCACACC	TATCCTGAAA	GCGTTGCATT	CTTTATGATA	GGTGTGACGA	TGTCTTTACC	420
ATTCCCACGG	CTGCTTTGCG	TGATGATGAC	ATTCATCATG	TATTTCCATT	CACACATACC	480
TTTTGTGCAT	ACGGTTTATA	TATGACCATC	CACGCTTATA	ACGAACCTAA	CAGTTTATTA	540
GCCCTTGACA	GGATAGGTCA	AAAGATTATA	TGTAGGTTTT	CCGGTAAACC	GAATTGTGAT	600

-continued

ATTTCTCTGC	AGGAAATAGA	ACAGCCTGGT	ACCTATAAAA	CGGACAATGC	AGTACTGTAG	6 6 0
CAGCGTAACC	AAGTAGGTCC	ACATGAACAC	GTACAAAATT	ATGGTAAGCC	ATCGTTTTTC	7 2 0
ATACCACAGC	CTGTAGCTGT	CGTACATGAA	TGAGGACGGT	CGAGGAACCC	AGGGTAGTTG	7 8 0
TAATTGGGGG	CGACATTCGT	ACTGTCCAGA	AGACAATTGC	ACGGGTTTCA	GTGAGATGAG	8 4 0
TACTTTAGCG	ATGTCGGCGG	GGGCGCTACG	TTTCACCGTG	ACGGTGAGAA	CTTGACCGTC	9 0 0
GTTTTGTATT	TCATGAGGCA	CGTTATACAA	GCCACTGGTA	TCATGAAGGA	TGACCTCTGA	9 6 0
TGCGATGTGA	GGATTAAATT	GTCCCTCAAA	CCGCCAAACG	CTGGTCATGT	TTCCACCGTC	1 0 2 0
AATTACGCAG	CTGACGGTGT	GAGATACCAC	GATGTTGGAC	TTAGGTTTGG	GGGCTAATTG	1 0 8 0
CCTTTTTACA	AATTCCCTTC	TGTATTGCAG	GTCCTGCTGC	CACTGCTTTT	CCGTGCGGAA	1 1 4 0
AGTCGCCATG	TCTTCCACAC	GTGTGGCGAC	GATAGACGCC	ACCAAGGTAG	CTACCAGAAG	1 2 0 0
CAGCTGGATC	CGCATGGCAT	TACCGTATGT	CAATTAGAAA	GTTGAGCGGA	CACGGTTATC	1 2 6 0
GTTCTGGCG	GATATAAGTA	TATAAACGCG	AGTTAGCCTT	TCCCGTCCGT	TTTGTACACC	1 3 2 0
CGTTCCCCAC	ACAAATGACG	AATACGACCT	TTTTTTTTAT	AAAAATAAAC	CACGTGTATT	1 3 8 0
ATATAAAAAC	ATTTACATAG	AAAAGAGACA	CACGGATCAA	CATAAGGACT	TTTCACACTT	1 4 4 0
TTGGGGTACA	CAGGCGTGCC	ACCGCAGATA	GTAAGCGCTG	GATACACGGT	ACACAGTCCT	1 5 0 0
GGCCAGCACG	TATCCCAACA	GCAGCACCAT	CGCCATACAG	ATGGCGATCA	CGACCCCGAG	1 5 6 0
CTCTAAGTGT	CTGTATTCAT	AGTGTAGTCG	CCGCAGGTTA	TCCACTGAAT	TCCCGTAACT	1 6 2 0
GAAATAACGT	ATATGGTACC	GAGGCTGGCA	CCACATGGGT	TTGCATTTGG	TGCACGGCAC	1 6 8 0
CAAATGCAGA	GTGAGATGGT	CCAAGTCCGT	GGGCACCCAC	TGGCGCAAAC	GGAATACGGC	1 7 4 0
TTCGGTGGTC	TCCACGAGGC	ACTCCGGGGC	GTGCAGACGG	CCCCACTTTC	GTCCGCGACG	1 8 0 0
GCCCGACCAG	CCGACCCGAG	CCACTATCCC	TTTCTCGGGA	TAGAACGTAC	CCTGTACACG	1 8 6 0
CCACACAGCG	TCCAACACGC	CGTCCTTGAC	GACGCAGCTG	GCCTGATAGC	TGGACACGTT	1 9 2 0
GTTAAGCGGC	GGAAAGCGAA	ACTGACGTGC	CGGCGGAGCC	ACATAGTTCG	GTTACACGTG	1 9 8 0
TTGTCGCGGT	TCGTCCTCCC	TATAGTAATA	GTAGTCGTCG	TCCTCATAGG	GGTTGCCGGC	2 0 4 0
GTGAGCCAGC	GTTACCCAAC	AGCAGCCCAG	GCCGACGAGG	AGGCGCAGCC	ACCGCCTCAT	2 1 0 0
GGCGGCTTCG	CCAGTCAATC	GTCTTTAGCC	TCTTCTTCCC	GTGAGGTCCT	TCCGGTGGCG	2 1 6 0
CGGTGCCGAC	CTCGGACCCA	GGGACGTATC	CACCTCAGGT	ACACACAGCA	GGCTACCTGG	2 2 2 0
ACACCGAAGC	TGAACAAGGC	TACGTGTTTC	ACAAACTGCA	CCAGTACCAC	ATAGAGGAAT	2 2 8 0
GTCAGGTAGC	GTCTCTCCGC	AAACAGCCGT	TCCAAGTCTG	AGGGCGTTAC	CCGCAGCGGC	2 3 4 0
AACCAGGGCA	GCCTGGACGC	CGGCCGGCAA	TGGAGCACGC	TCCGGTTACA	GGCACTGCAG	2 4 0 0
GGGTAAACGG	TTAACATCAC	GTAAGAGAGT	CGTGCGTCCA	CCTGTGGGAG	CTCAGTTTCG	2 4 6 0
TAACGTAGAG	CCCCGTCATT	TTCCAGCTGG	GGTGCGCCGA	CCTTGAAATG	GGTCGCGCTC	2 5 2 0
CGCTCGTTAC	CCCAGGTGCC	GTAGGCTCTC	GGGGCCGTAT	CGGAGAAGTT	GCCACGCACA	2 5 8 0
AGCCAGGCGG	CCACGAGTAC	CCCGTGCTGG	ACGTAACATT	CGGACACGGA	ACTGGAGACA	2 6 4 0
CGGTAGCCGG	ACACGTCCCC	AAACCCGCGA	GGGTACTGGG	GCAGACGGAC	GGACTTGCTA	2 7 0 0
TTTGACAACG	GACAGATACG	AGACGACGAG	GACGCAGACG	ACTCGTCGCT	GGACCACGAC	2 7 6 0
AACCGGAGCG	ACTCCTTGGA	GCGGCTCGAG	AGTACACTTA	CTGCGATCAG	ACACCAGTGC	2 8 2 0
CAGAAGAAGG	AACAGGTGGA	CGGGGACCAC	AGGATCATAG	CCGCCGGCAC	CGCGGCCGGC	2 8 8 0
CGCAGGAAGC	CGCCCGGCGC	GTCGTCTGTG	TGCGGGAGCC	GAAACACCGT	GCCTCTTTAT	2 9 4 0
ATCGTCCCGA	CGTGACGCGA	GTATTACGTG	TCAGGGGAAA	CCCCCGTCAC	GACGAACGTG	3 0 0 0

-continued

ATTTGTAAGT	GACGCGGGGT	GCTGACGGGG	TTCGGCCCGA	GAGGTGACGG	AGCGCCTCAC	3 0 6 0
GTCAGTATGA	TGTCCGATCC	GCGTCAGCCC	CGACGTGGTT	GTGGTCACCG	AAACCCACGT	3 1 2 0
TTATATGGAC	GTTGAGAGCA	GCGCCTGACC	ACATGATTCA	TCATACCATT	TCTCGGAATC	3 1 8 0
GGGCCCATGC	CGGGAAAGCA	CATTCCTTTT	CAGTAAACAA	CAATGACATC	ATAACAAATC	3 2 4 0
ATTTTATTCG	CGAGGTGGAT	AATAACCGCA	TATCAGGAGG	AGGGATCGGG	TGATGACGCA	3 3 0 0
GGCCCCGCAG	AACAGTCCGA	AATAAATTTT	TAGTATTGCC	CCATAGTCGC	CTAGATACCA	3 3 6 0
GAGGTACGTT	AAGTTCATCA	AAACGCCCAT	CGGCGTCCCG	GAATCGTATA	CCGGGCACAC	3 4 2 0
GAAGCGTTCA	TAACAATCCC	GGGAGGCGAG	TGTTAGGGTA	GCAGAGTAGT	TTCGGGGTCG	3 4 8 0
GTTTCCTTCC	GGCGACGACA	GTTCCGTGGG	CAGCAGAATG	TACAGCGCCT	CGGTAGCTGT	3 5 4 0
CGCGGTGCCT	TCCACGAGGA	TGGGCTGCCG	GTGCCTTTTCG	TGATTTTCCC	CGTCGTGTAG	3 6 0 0
CCAAGCCGAG	GCCCGCAAAG	TCTTAGGCCG	GGGGAATTGT	CCATAGAGTT	TCACCGCACC	3 6 6 0
CTTCAGTACA	TGGTTCTGAA	TAACACAGCC	GCACGTGAAG	TAGGTAGGTT	CTCTCGTCTC	3 7 2 0
CTCCGTGGCT	GCCGCCACCA	CTCCAGCCA	CCACAACAGG	CAGATCGCCA	GAGGGTTCCG	3 7 8 0
GAGGCTTCCC	CGGCGTAGCA	TGGTTTTGGG	TTAAAGCAAA	AAGTCTGGTG	AGTCGTTTCC	3 8 4 0
GAGCGACTCG	AGATGCACTC	CGCTTCAGTC	TATATATCAC	CACTGGTCCG	AAAACATCCA	3 9 0 0
GGGAAAATGT	CGGTGCAGCC	AACCTTTTAC	ATACAGCCCC	CAAAACACTT	GAATCACTGC	3 9 6 0
CACCATCATC	AGCGTATACT	GCGCCGACTT	AATCGTGAGC	GCGTAGTACG	CCATTAGACG	4 0 2 0
GCGATCTTCG	AACAATAGTC	GTTTCGATGTC	CTCTAACGAG	CTCCACAGGG	GAACCCAAGG	4 0 8 0
CACGAGGCAC	CGGGGTTCGC	ACTCTACATA	ATAAGTTTGG	CATTGGTGGC	AGGGGGAAAA	4 1 4 0
GTAGAACAAAC	ACGAGTTTTG	TGCGTTGGGG	AACACGATAG	TCCCGGAGCC	AGTAGCGTTT	4 2 0 0
TGCGACGAGG	CTTTCGGAGA	CGTCCTCCAC	CGGCGTCGGC	ACTCGATCCG	CGTAGCCCTC	4 2 6 0
CAGCGTCTGG	TAGTACACCC	GGGGTGTCCG	CGTGGGCACG	GACAGGTTC	CGCGCAGGGT	4 3 2 0
CCACAGAGCC	TCCAGTCGAC	CGCCGATCG	GAGCACGCAG	CGCGCCTCGG	AATACTCTAC	4 3 8 0
TCGGTACTCC	GAAACATCGG	ACAGAGGCGG	TAACGGCTCC	GTCTCCACCA	AGGGCGGAGG	4 4 4 0
TTCATCGAAA	AGAGTCAAGG	ATAATTCAGG	CATACTACCC	GCGACCGGGG	CCCAGAGGGC	4 5 0 0
TAGAATAAGC	ATTACAAGGT	TCATTCTGTC	TTACAAGGGA	AGGCTGTTAC	CCTGTCTAGA	4 5 6 0
CTCAAAAGCT	GTAAGGCTGT	CTTATAGCAT	GTAGTCTTGC	ACGTCACGGG	GAACAGGGTG	4 6 2 0
GTGATCTAGT	GACGTCGGGA	GAACACGGTG	TTTTAGGGTG	CGGGGGACAA	AGGACAGTAC	4 6 8 0
GACAGATTAG	GTGATAGAAA	CGTTTTTTTT	TATTTATGAA	AAAGCCAGTG	TGCCGTGCGG	4 7 4 0
CCTAGGGCCC	CGGCGTAGTT	TGGATACCAG	ATGGGGGCCG	TCAGGGGTAC	TACCACGAGC	4 8 0 0
AGAAACATAA	TGACTTGGTC	CATGTATAGC	AGCATAGCGG	TGCGCAGCAG	GTCGCCGTCC	4 8 6 0
GTGTAGCAAT	TTGACGGTGA	GCGATAAAGC	ACCGTTAATG	TGTCGCGGAT	AAGCACGATC	4 9 2 0
TTGAGGCCGT	AGATGAAGCT	CACAGTCAGT	GCTAAAATGA	TGCGTTGGTA	TGGTTCCCAG	4 9 8 0
GACTGCACGG	CGATGAAGAG	CCAGAGTATG	GGAAGCATGA			5 0 2 0

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5924 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

-continued

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCTAGACTCA	AAAGCTGTAA	GGCTGTCTTA	TAGCATGTAG	TCTTGCACGT	CACGGGGAAC	60
AGGGTGGTGA	TCTAGTGACG	TCGGGAGAAC	ACGGTGTTTT	AGGGTGCGGG	GGACAAAGGA	120
CAGTACGACA	GATTAGGTGA	TAGAAACGTT	TTTTTTTATT	TATGAAAAAG	CCAGTGTGCC	180
GTGCGGCCTA	GGGCCCCGGC	GTAGTTTGGG	TACCAGATGG	GGGCCGTCAG	GGGTACTION	240
ACGAGCAGAA	ACATAATGAC	TTGGTCCATG	TATAGCAGCA	TAGCGGTGCG	CAGCAGGTCG	300
CCGTCCGTGT	AGCAATTTGA	CGGTGAGCGA	TAAAGCACCG	TTAATGTGTC	GCGGATAAGC	360
ACGATCTTGA	GGCCGTAGAT	GAAGCTCACA	GTCAGTGCTA	AAATGATGCG	TTGGTATGGT	420
TCCCAGGACT	GCACGGCGAT	GAAGAGCCAG	AGTATGGGAA	GCATGAAGCT	TAGCAAACAG	480
AGGATGGCTA	ACCGTCGTTG	CATGTTCCAG	GCCATGAGCC	AGGCTAGGCC	CGTACACCAG	540
ACGCAGAGCA	TGGATGACAG	GACATAGGCC	TGGATTACCA	CGGTGCGATC	GAAACACAGC	600
CCGATGGTGG	ACACGGATAT	CGTAGTGAGG	GTGGTATATA	CCATGACCAG	CATCAGGGTC	660
CCGGGTCGGC	GCCGACGTTC	CAGCCAGTAC	GCGTGGCAAC	GCAGAGCGCA	GGGTAGCAGT	720
GTGCTCCAGA	AGGGCAATGT	ATCGCGCAGG	TAGGGGGCCG	TCACGCGCCA	CGGTATGAGC	780
ATGAAAAGGA	TGGTAGTGGC	TATGGTGGCG	CTGGTCTGGA	ACACGACAGT	GCCGTAGAGA	840
CGTACCATCC	AGAGAAAGTG	TTGAACGCTC	CGCAGGGTGT	CTTCATCTTT	GGTGATTACG	900
GTGACTCGAC	GGATCGGCGG	TGGTGACGGC	GGCGACACGG	GTGGGGGTTT	CTCTTTCTTA	960
TGGCCGAGTG	GCTCGCCTTG	GTGAAACTGG	ATCTGTACCA	TGACGGGTGC	TCGACGAACA	1020
GTCGTGGGGG	CTTTAGGTAC	CCGGCAAGTT	TTATAGAGAA	AGGGGGACGA	TGGGTGGTGG	1080
CTACGAGCCA	CCGCCACCTT	CGCAATACGA	GGATCTGAAG	GCGGCAAAGA	CGGTCGTCCA	1140
GGGCAGGCGC	CAGAGGTTGG	GACTGAGCAC	GATCAGCGTG	ATTTTAAACA	TGGTCACCAG	1200
TCCTACGTAG	ATCAGCAGCG	AGCCGCGTAA	CGTCTGAGCA	GCCGGCAGTT	CGTCGCGGAT	1260
GTAACGCGTG	CCGTAGAAAG	TCACGGTCAT	CATAAGGAAG	ACGATGGCGC	CGTAGCCGTA	1320
GAGTAGAATA	CGCTGATGAT	GGAACACGGT	CTGGTCGCCG	ATAACCCAGA	GCGTGATGAA	1380
AAAAACGCTG	GTGAGTACCC	GTGAGCATAT	GAGCTCCCAA	CGCTTAGCGC	GAAAGCTGTC	1440
CCCAACCATG	ACAGCGCCGG	TGCAAGCTAT	CCACAGCGTG	AGGACCAGTG	TGTAGTCGAT	1500
GAGGATGGCG	GGCAGGTCGG	AGCACCAGGT	GTAGAAAACC	GTGGTAACGG	AGAGGAGGCC	1560
TACGTAGCCC	ATGGTCAATA	CCACGTGCTC	GGGGTGCCCT	TCGCCCTGTA	TCAAGACCAA	1620
ACACCAGAGA	AGGGAGGGGG	CAAAAACCAG	CAGCAGAGGG	GAAGATTCAT	GTTGACATAT	1680
GTTGTGGGAA	TCGGGGATAC	CCAGCCAAAT	CATTCCGCAG	AAAGCCGTAC	TGATGGCGAT	1740
GTGAAAGACC	ACTAGGGCGT	AGACCCGGAC	GAGGACAGCA	AAACGGCGCA	GCCACATAAG	1800
GCCGTGGTGC	AGCTGCAGGA	GGGAAGCCCA	TTGCGGCGAA	TGTAGCGACG	GTAGCGGCGG	1860
GTCCATGAGG	CGGGTGATGC	GCCCGAGTGA	ACGGGTGAGC	GTCTCGGTGG	AGTCTTCTTA	1920
TAAACCAGCG	GAGCTCAGGC	AGCCTTGCTC	TGGAACGTCG	CAGTGGTGGT	GTTGAGGATG	1980
ACGCTGAGCG	TGCCGTTGTC	AATCAGGTAA	TGATGATAGG	TGCCGAGCTT	GGCCAGGTAG	2040
CTGAACATTT	GGTCCCAGCG	TGCCGACCAC	ACCACGGGCG	TGAGCATCAG	GAGTGTGGTG	2100
TGATAGATTA	GTGTTTCGGT	GGCGTAAAGT	ATCAGCGAGC	TGCGGATGAC	GTGGCTCACG	2160
GGCATTTTGG	TGGCGATGTA	GCGCACGTCT	TGGAAAAGGA	CGGCCAGGAT	GCAGCCCACG	2220
AACACGGTGT	AGAGACACAG	CAAAGTCTTA	TGTAACCAGG	TGTAAGTAGA	AGCCAGGACG	2280
CTGACCATCA	CCGTCAAAAAG	TGTGGAGGTA	AAAAGCGCGT	CACGCCACAC	GGAGCTGAGA	2340

-continued

CGGTGCTCCC	AAGCCACGCC	GTTGCAGGCC	ACGAACAACG	TCCACGTTAG	GATGAGGCTA	2 4 0 0
GAAATGCCGA	TGGGCGCTGT	GGCGCACAGG	TTGAGCCCGG	CGGTGGTGAA	CGAGAGAAGC	2 4 6 0
GCCACATACA	GCGCAAACAC	CAGGCCGTTG	CTGGGGTGTC	TGTGATCGGT	GAGCTCCAGC	2 5 2 0
GCGCCAGAA	CCAATACTGG	TGTGCAGCTA	AGCAATAGCG	GCGAGGGATC	GTCGCTGCAC	2 5 8 0
TTGTAGCCCA	GCGAGGGGTA	ACCCAGCCAA	ACCAGCGCGC	TAATGAGTAC	GCTGAAAGCG	2 6 4 0
GTTTCCAGCG	TCAGCAATCC	GTAGACACGC	ATGACAATCG	CGGTCCGCCG	TAGCCAACAC	2 7 0 0
ACGGCATCTT	CGGAAACTGT	GGACGCTGTT	TCCGAATACC	GGGAGGAGAT	CGTGCTTCCC	2 7 6 0
TCTTCCAAGG	ATCGGAAAGT	AGCGTCCGTC	GTTTCCGCGG	ACGCGGCTTC	CCTGGTACGC	2 8 2 0
TCCGTTTCCG	ACGACGCGGT	TTCCCGCTGC	GTGGAAACTG	TCTCCATGTC	GGGACCGCAG	2 8 8 0
CGCCCGGCGG	CGTATCCGCA	AGGTCTCGAA	GCTACAGCTT	GTCAGAGGAA	AAGTAGGTTT	2 9 4 0
GCAAAAAGGT	GCGCAGGGTC	ATGATTCTCA	GCACCATCAG	CAGAGTGAAA	ACCAGACTGA	3 0 0 0
GAAACACCTT	GACGGCCGCC	AAAAGCGCGC	GTTCCAGCGG	CGTCTCGTAG	CGTACAGCCA	3 0 6 0
GGGCCGCTTC	GTGGAAATGC	GAGACGGCTA	GACAGGTAAT	GAGCACGCTG	AAGGACAAGA	3 1 2 0
CGATCTTAAA	GCACCAGGAC	CAACCACGCC	TCAAGATGAC	CACCACGATT	GCCGTGAAGG	3 1 8 0
TCAACGTGAT	CAAAGCATGG	ACGACCACGA	TCTGACGGCG	GACGGTACGT	TCGGGAGCCA	3 2 4 0
ACAACGCTAC	GCCGGTGCAG	CTGAGAAAGG	CCAGTAAGGT	GAACAACGCG	GCCGAGATGA	3 3 0 0
CCAACGTACC	GTCAGGCAG	AGACATATCA	CGATCAACGG	CGGCACGTGA	AGCAGCGTGT	3 3 6 0
AAAAGAGCAG	AACGCCGATA	TTGCTGGGAT	GCGATGTTTC	GTAACAGTGA	ATGAAGATCA	3 4 2 0
CTGACGTGAC	GGGTATGACA	AAGACGAGGC	TGGGCGAGGA	CTCCGTGAGA	CACAGACGAG	3 4 8 0
AATGGTGAAA	CCACGTCGCG	GGCGCCGCGT	AGCAGAAGGC	GCTCAACAAC	GCGGTCAAGC	3 5 4 0
CGGCCAGCTG	CCAACCCACG	GCGCCATAGG	TGTGCAGCGC	CACGCGGCAA	CAGTCGACCC	3 6 0 0
AAGCCAGACT	GCGGGTCGCC	AGCCGGGTCT	CTTGGATCCC	GGGGGGCACG	TAGATGACCG	3 6 6 0
TGCCATCGGT	GGGTACTTGA	AACCCTTTTT	CTCTTCTCAT	GGTGCCTGTC	GTTCTCTGGA	3 7 2 0
AACGGCTGCT	CTGTCCGAAA	ACCAGTTCCG	AACGAAAATC	TAGGGCGAGA	GGGTGGACAA	3 7 8 0
CGGCGTCGAC	GACGAAGCAT	GGGACAGGTC	GTTCCGGCGTT	AACGTCATCG	CGTCGGACGA	3 8 4 0
CGGTAGTTCT	AAGAGACGTA	GATCGCTCAG	CAGGTCCTGA	CAGTTGCGGA	TTCGCAAGAT	3 9 0 0
CAGAAAAAAA	AGGGAAATGA	ACGTAATAAA	GAGCTGTAGC	GACGTATGCG	CCACATCGCG	3 9 6 0
TGGCATAAGA	ACGTGACGGA	CGAAAAGGAC	CTGCTGCGAA	AAGTGACCGG	CGAAGATAAG	4 0 2 0
GCCCACCGTG	CTGTAGAAGC	CCAAAAGCAG	CCGCAGGGGC	CAAGTCCAGG	GCCGCGTGAA	4 0 8 0
GACGATGAGA	ACGTTGACCA	GAAAGACCAC	GACCCAGACG	CCGTTGATGA	GGGTAAATTG	4 1 4 0
ATCGGACAGG	GTGCAGTTGT	CGCGACAGAT	GAAGACTACT	TCCGCGCAGA	GCAAGGTGAT	4 2 0 0
GACCAACGTG	AGCACAAACG	ACGTCAACAC	CTCGCGGGGC	TCCTGGCAGG	CACACGTGAC	4 2 6 0
ACCTAGCGCC	GGGATGTGCG	CCAGGAGGCC	GGCGAGTAAT	AGCACCAGCT	GTCGGAACGG	4 3 2 0
ACGACGGCAG	CGCGGGTGCC	GGTTTCGCTG	AGCGAGAACC	GGTCGCTCAT	AGCGGAAATA	4 3 8 0
CACGAAGAGC	GCGGAGGCCA	CAGGCACCAG	GAGGAGCACC	TCGGGCGCCC	AGACAACGTG	4 4 4 0
ACAAGGAAAG	CCCGGACGCG	ACTTGAGAGT	CGCTGTAGGG	AAGACCAGAG	AGAAGCTACC	4 5 0 0
CAAGACGGCC	ACCGCCGCGG	AGATTTGGAA	GAGGAGCAAG	CCGGCGATTC	GGACGACAAC	4 5 6 0
CTCGAAGCGA	TGCACCCAGC	CCAGCACGGC	CACCACGGCC	GCTTCATCAT	AGTCGTCTGT	4 6 2 0
GTTGCCGCTG	TCGAACAGCC	GCCGAAACAC	GATCTGTGCG	TGGGTCGCGG	TGGGAAAGCG	4 6 8 0
CAGACCCATG	ACAGCCGGAG	GCTATATGAC	CGCGCGTCTA	AGACGCGAGA	TCCGTGGGGG	4 7 4 0

-continued

GACTTTTAGA	TGTTTGGGCG	GCCCGCGGTT	CTAACAGGCT	TGATTGGTGG	AGACGGCCGG	4 8 0 0
CGCGGCGGGT	GGGGGAAACG	ACGAGTTTTT	CCGTTACGCC	ATGGTTCGCG	TGAGGTTTCT	4 8 6 0
CTGTACCTCC	CGCAAAAGGT	CACAGCCCGA	AATGGAGGCC	GCGTTGGTGG	CCCCGGTGGC	4 9 2 0
GCGTGACGAT	AACCAGGTCA	TCCAAGCGAT	GAGTTTGTCT	AATGAGTCCT	CGGTGGTGAA	4 9 8 0
GAGGATGAGA	ATGAGCAGGT	ACAGGTACAC	CAGGTTCTCA	TAGAGACACA	AGGTGAGCAG	5 0 4 0
GTCAGCCTCG	GACCACGCGA	TCTCAAACAG	GCGCGTGGTG	TCAAAGACCG	TGACGACCAG	5 1 0 0
CATGAAGCTG	AGCGCCATGG	CGTAATAGCC	CAAAAAAAGT	TTGTGCCCCA	ACGGTACGGG	5 1 6 0
CTGCAGGTAA	AGTGCGATCA	AGAACGCGAT	AACGCCGATC	ACAAACAGCG	TGACGATGAC	5 2 2 0
CTGCCATCGA	CGGTGATTAT	GGCCGGCTAG	ACCCGTGACG	CAGCTGCAGA	GGCTAAAAAG	5 2 8 0
CACGCAAGCC	AAGAGGCCCG	AGAAGGTCAC	TAGCGTAGAG	GAGGAGCAGG	CGCTGGCCAC	5 3 4 0
GATCACCGAA	AGCGTCGTGA	GCACGCTATA	AATGGTGAGC	AGGCCAGGGC	TCGGTGGCGA	5 4 0 0
CGTGAACGAT	CCTTCATCGC	GTTTGCCGTG	CAGCAGGGCC	AAACAGATGG	TGGGCACCAT	5 4 6 0
CAAACCTAAG	GGCGGCATAA	AGCCGGTGCA	ACAGAGAAAG	ACGGTGCCTT	TAAGATGCGG	5 5 2 0
AAAAGCCAGC	ACCAGGCCCA	GACAGAGCAA	GAAGGTGCAG	GTGCCCTGCA	CGGCCACGGT	5 5 8 0
GCTGTAGACC	CGCATACAAA	GTAAAAAGCG	ACGTACGTCG	TTCGTGACCA	CGGAGGAAAT	5 6 4 0
CATAATGACT	CCGCGCGAGG	GTCGCGGGGG	TGGGGGCGCC	CAGGCCGTCC	CGGTGGCCTC	5 7 0 0
TGAGTTCGGA	GACATGACGG	CGGTGGCGAT	CAAAAGGCGC	GTATGAGAAA	CCGTTTATAG	5 7 6 0
AGTGTAATAG	AATCACCGTC	ATTCCACAC	GGCGTTCCCC	CATAAAGTCA	CGTAACACTC	5 8 2 0
GAGTAAGCGT	GAAAAAGCTT	TATTGTTGAA	TAAAAAACAC	GAGTACAACA	CCGAGTTGCG	5 8 8 0
GTGTCCTGTC	TGTCTACTGG	GTGGGGAAGG	TTCATCGTCT	GTCT		5 9 2 4

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1707 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGCAGTGCTT	CGCTGAGAAG	TAGCGTGTGG	ACTGAACGGT	GTTTTTGAAT	ATATAGCGTT	6 0
TCTTGGTGAC	GTTGTTTCCC	CTACGTAGTA	GGCAACTACG	TGCCAAAAGA	GGCGTTACGG	1 2 0
TACTTTCCGT	ACTGGGATTT	CCAAACCGGG	ACTTTCCACA	CGGCGGTTTC	AACACCGGGA	1 8 0
CTTTTCACAC	GGTGATTTTCG	GCACCGGGAC	TTTCCGCACG	GCGGTTTCGC	CACCGCTGAC	2 4 0
GTTCTCATCG	CCGCCACGT	CAACGGTGGC	GACACCGTAC	TTTCCCATGC	GGTTTATAAA	3 0 0
CGTCAAGAGT	CACGTCAGTC	GCCCACCCCC	ATTACACGGC	GATATCCCGA	TAGGGCATGA	3 6 0
GGGGACCCGG	GTGTCGCGAC	ATGTCGACGA	CAGGTGCGGA	TAGTGGTTCG	TGTCGCGACA	4 2 0
TGGACGTGCA	GGGGGATGTC	TGTCGCGATA	GAGTTGATGT	GACAGCCCGC	TACACCTCTC	4 8 0
TGTCGCGACA	TGCATACACA	ACGGGCCGGC	TTGTCGGCGA	TTGTCGCGAC	ATATCGTTAT	5 4 0
CAGTTAGCGA	CCGGAGTTGT	CTATCGCGAC	ATATCGTCGA	CTATCGCGAC	AGAAAAATA	6 0 0
CCGTTTCGTAG	AGAATGCCGT	GTTGAAGGAA	CGCGCTTTTA	TTGAGACGAT	AAAACAGCAT	6 6 0
CAGGAGCCAC	AACGTCGAAT	CCCACGTCCA	GTCGATTCGT	ATGTTATGCT	GCACAGCAAT	7 2 0
GCTAGAATAA	CAACCAGCAG	GGTAATCCCG	CAACATAAAT	ACAAAGTCAC	AGCGAAGAAT	7 8 0
CCGTGTCGTT	CTATCAAGCG	AAACGCGTTC	CAAACGGCCC	CGTCACAGAC	GCAGTTATTC	8 4 0

-continued

ATAAGCGTTA	ACAACCGGTG	GCTAGGATGA	ATATCCAAAT	CACAGGGCAG	TAGCCGACGG	9 0 0
ACTCGTTGAC	AGGTCAGCCT	ACCCTCAAGG	TTCCTATCGT	TCGGACGGGA	TTTGTGCGTT	9 6 0
TTAGGCCTCT	TTTTCGCCGC	CTGCAAGCAT	TGGTGCGCAA	AGTCCTCACC	CAGCTGTTTC	1 0 2 0
CAGCTATCAT	CTGCATCTGT	GCAGTCCCCT	GTATCGTTGT	AACAAACGGG	TCTGTGCGAC	1 0 8 0
TTCGTTCTCG	GAACACAAGC	TTGTTGTTCG	GGAGACAGAG	AGAGAAGGGT	TTTCGGGTCA	1 1 4 0
CGCGAAGACC	GCTCACCGGG	GGTCGGCAAC	GCACACATCA	ACAGAAAACC	GAGACGAATC	1 2 0 0
AAGAGATCCA	TAGTGAAGGA	GTGATATCGA	CGTGCTTACG	AAACGGCGAT	TATATATGTT	1 2 6 0
CTCAACAATA	CCGCCCTACG	TTGTATGATG	TAACGTGTGA	CGTGAGTCTG	ATCCAACACT	1 3 2 0
GAACGCTTTC	GTCGTGTTTT	TCATGCAGCT	TTTACAGACC	ATGACAAGCC	TGACGAGAGC	1 3 8 0
GTTTCATCGGG	GCATGAAGTA	CGCATTACAC	AAACTCCATA	TATTTGTTAC	GATAGAATAC	1 4 4 0
GGAACGGAGG	AGGCTTTCGC	CACACCTATC	CTGAAAGCGT	TGCATTCTTT	ATGATAGGTG	1 5 0 0
TGACGATGTC	TTTACCATTTC	CCACGGCTGC	TTTGCGTGAT	GATGACATTC	ATCATGTATT	1 5 6 0
TCCATTCAACA	CATACCTTTT	GTGCATACGG	TTTATATATG	ACCATCCACG	CTTATAACGA	1 6 2 0
ACCTAACAGT	TTATTAGCCC	TTGACAGGAT	AGGTCAAAAG	ATTATATGTA	GGTTTTCCGG	1 6 8 0
TAAACCGAAT	TGTGATATTT	CTCTGCA				1 7 0 7

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1817 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCTAGACTCA	AAAGCTGTAA	GGCTGTCTTA	TAGCATGTAG	TCTTGACAGT	CACGGGGAAC	6 0
AGGGTGGTGA	TCTAGTGACG	TCGGGAGAAC	ACGGTGTTTT	AGGGTGCGGG	GGACAAAGGA	1 2 0
CAGTACGACA	GATTAGGTGA	TAGAAACGTT	TTTTTTTATT	TATGAAAAAG	CCAGTGTGCC	1 8 0
GTGCGGCCTA	GGGCCCCGGC	GTAGTTTGGA	TACCAGATGG	GGGCCGTCAG	GGGTACTION	2 4 0
ACGAGCAGAA	ACATAATGAC	TTGGTCCATG	TATAGCAGCA	TAGCGGTGCG	CAGCAGGTGCG	3 0 0
CCGTCCGTGT	AGCAATTTGA	CGGTGAGCGA	TAAAGCACCG	TTAATGTGTC	GCGGATAAGC	3 6 0
ACGATCTTGA	GGCCGTAGAT	GAAGCTCACA	GTCAGTGCTA	AAATGATGCG	TTGGTATGGT	4 2 0
TCCCAGGACT	GCACGGCGAT	GAAGAGCCAG	AGTATGGGAA	GCATGAAGCT	TAGCAAACAG	4 8 0
AGGATGGCTA	ACCGTCGTTG	CATGTTCCAG	GCCATGAGCC	AGGCTAGGCC	CGTACACCAG	5 4 0
ACGCAGAGCA	TGGATGACAG	GACATAGGCC	TGGATTACCA	CGGTGCGATC	GAAACACAGC	6 0 0
CCGATGGTGG	ACACGGATAT	CGTAGTGAGG	GTGGTATATA	CCATGACCAG	CATCAGGGTC	6 6 0
CCGGGTCGGC	GCCGACGTTC	CAGCCAGTAC	GCGTGGCAAC	GCAGAGCGCA	GGGTAGCAGT	7 2 0
GTGCTCCAGA	AGGGCAATGT	ATCGCGCAGG	TAGGGGGCCG	TCACGCGCCA	CGGTATGAGC	7 8 0
ATGAAAAGGA	TGGTAGTGCC	TATGGTGGCG	CTGGTCTGGA	ACACGACAGT	GCCGTAGAGA	8 4 0
CGTACCATCC	AGAGAAAGTG	TTGAACGCTC	CGCAGGGTGT	CTTCATCTTT	GGTGATTACG	9 0 0
GTGACTCGAC	GGATCGGCGG	TGGTGACGGC	GGCGACACGG	GTGGGGGTTT	CTCTTTCTTA	9 6 0
TGGCCGAGTG	GCTCGCCTTG	GTGAAACTGG	ATCTGTACCA	TGACGGGTGC	TCGACGAACA	1 0 2 0
GTCGTGGGGG	CTTTAGGTAC	CCGGCAAGTT	TTATAGAGAA	AGGGGGACGA	TGGGTGGTGG	1 0 8 0

-continued

CTACGAGCCA	CCGCCACCTT	CGCAATACGA	GGATCTGAAG	GCGGCAAAGA	CGGTCTGTTCA	1140
GGGCAGGCGC	CAGAGGTTGG	GACTGAGCAC	GATCAGCGTG	ATTTTAAACA	TGGTCACCAG	1200
TCCTACGTAG	ATCAGCAGCG	AGCCGCGTAA	CGTCTGAGCA	GCCGGCAGTT	CGTCGCGGAT	1260
GTAACGCGTG	CCGTAGAAAG	TCACGGTCAT	CATAAGGAAG	ACGATGGCGC	CGTAGCCGTA	1320
GAGTAGAATA	CGCTGATGAT	GGAACACGGT	CTGGTCGCCG	ATAACCCAGA	GCGTGATGAA	1380
AAAAACGCTG	GTGAGTACCC	GTGAGCATAT	GAGCTCCCAA	CGCTTAGCGC	GAAAGCTGTC	1440
CCCAACCATG	ACAGCGCCGG	TGCAAGCTAT	CCACAGCGTG	AGGACCAGTG	TGTAGTCGAT	1500
GAGGATGGCG	GGCAGGTCGG	AGCACCCAGG	GTAGAAAACC	GTGGTAACGG	AGAGGAGGCC	1560
TACGTAGCCC	ATGGTCAATA	CCACGTCGTC	GGGGTGCCTT	TCGCCCTGTA	TCAAGACCAA	1620
ACACCAGAGA	AGGGAGGGGG	CAAAAACCAG	CAGCAGAGGG	GAAGATTCAT	GTTGACATAT	1680
GTTGTGGGAA	TCGGGGATAC	CCAGCCAAAT	CATTCCGCAG	AAAGCCGTAC	TGATGGCGAT	1740
GTGAAAGACC	ACTAGGGCGT	AGACCCGGAC	GAGGACAGCA	AAACGGCGCA	GCCACATAAG	1800
GCCGTGGTGC	AGCTGCA					1817

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1702 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCATGGTTCC	GAAGCGTCCC	CACATGCACC	AGCAGTCGGC	GTCAAAGTCG	CTTGCGCTGT	60
CGGCCCAGTC	GCCACCGCCG	CGGCGGATTT	CCGCGCGGGG	GACGGGGTAG	CCGAGTGCTG	120
CGCCCTCGCC	AATGTTGTGA	AGTGGATGCG	TGAGTTGATG	TTGATTCTCT	GTGGGAAAAT	180
GAGCGCTGTC	CTGTGGGTTG	GTGTTGGGGT	ATGCGAGTAG	TAGGGGTTGT	GTTTGATCGT	240
AGAGGTGTTG	GCGGGCCTGT	GCGCAAGCAG	CGTAGTCTGC	GGCGTCGAGC	TCCATCTGTG	300
TGCGGTGTTT	TTCGTCGGCG	TGTTTGTCCG	AGGTTTGGAC	ATGCGGTTGT	GTGTTGCTGT	360
GGTGTAAAGG	TAACGTGTGT	TGGGCGTCTG	GGTGAAGCGG	CGTGGTGTGG	GTGCTGTTTG	420
TGTCTGTGGC	TGGCATGATT	GTGCGGCATG	TGTGTGTTGT	AGTGGGTGGA	GGTTAAATAG	480
GTGAGGTGGG	TTCCCTGGTC	CGCGCCGCAA	ACTGTCCCCG	TCCCCAACGT	AACCTCCCCT	540
ACGCGGCGCG	AACAGCCCCG	GCCCCAGCGC	AACCCCCGTC	CCCGGCCCCA	ACACCGTCCC	600
GCACACCCCC	CGTCTCCGCA	ACACCCCCGG	ATCGCCGGCG	GCCAGAACGC	TCGAAAACCC	660
CCGACAAGCG	CAGCGCCGAA	ACGACACAGG	CAAGGACCGT	GGAACGCACC	GGCAGCGCGC	720
CGAAACACCG	TCCCGAAGCC	CGGTGCCGAC	AACAAATACC	GTGGGACGAC	ACGCACCGGC	780
AGTGCGCAGG	CAGCGTCGGA	CACAACACGC	TTACGGCCCT	CAACACTCCC	TCGAGGACCC	840
ACCACGCGGC	CCCGCACCGG	CGGTGTTTTG	GGTGTGTCGG	GGCGCGGCCG	GGTGGGTGTG	900
TGCCGGGTGT	GTCGCGGGCG	TGTGTTGGGT	GTGTCGGGGG	TGTGTTGGCA	GGGTGTGTCA	960
GGGTGTGTCG	CGGGCGTGTG	CCGGGTGTGT	CGTGCCGGGT	GTGTCGCGGG	CGTGTGGCGG	1020
GTGTGCCGGC	GGGGTGTGGT	GGCGGGGTGT	GTCGGCGGTG	TGCGCGGCCT	CGGGGTGTGC	1080
GGCTTCGCAG	GAACGAGTGT	GTGGCCTCGC	GGCCGTTATT	TCCCCGCGG	TCCCCAGGGC	1140
CGTCGTCCCT	CGCCCCCGGG	CGTTGCTTTT	CGTGTGTCCC	CAGGGACCCA	TGCTGCCGTC	1200
CCCCGGGAAC	TTCCTCTTTT	CCCCGGGGAA	TCACACAGAC	ACAGACACGC	GTCTTCTTTT	1260

-continued

CGCCGTGCGC	GCCGCACGTC	GCTTTTATTC	GCCGTGCGCG	TCCTCCGCAC	CACACGCAAC	1320
TAGTCGCCGT	CCACACACGC	AACTCCAAGT	TTCACCCCCC	CGCTAAAAAC	ACCCCCCGC	1380
CCCTCGAGGA	CCCACCACGC	GGCCCAGGAA	GGATGTGCGG	CGTCCACCTA	GATGGGTGCG	1440
CGCCCAGGAG	GCGGCTGTGC	GCTCCAGTGG	TACGCGCCTG	CCGCGCGTCT	TCCTTCGGGT	1500
AGCTGCCTTT	CCCAGTCCAC	GGCCTTCCAG	ACTGCGTGGC	GCCAAGGCGG	CGCCAGCACG	1560
CGCCGTGCAC	GTCGCTGCCT	ATAAAAGCCA	GCTGCGTGTC	GCCCAGCGCA	CACGGGCGAC	1620
GAAGGCGTCC	GCGTGTCTAA	ACCGCGTGCT	CGCTGACGCG	GGTTTGCTTC	CTATATAGTG	1680
GACGTGCGGAG	GTGTCCGGCG	CC				1702

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 256 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCATGCCGGG	AAAGCACATT	CCTTTTCAGT	AAACAACAAT	GACATCATAA	CAAATCATTT	60
TATTCGCGAG	GTGGATAATA	ACCGCATATC	AGGAGGAGGG	ATCGGGTGAT	GACGCAGGCC	120
CCGCAGAACA	GTCCGAAATA	AATTTTTAGT	ATTGCCCCAT	AGTCGCCTAG	ATACCAGAGG	180
TACGTAAAGT	TCATCAAAAC	GCCCATCGGC	GTCCCAGGAA	CGTATACCGG	GCACACGAAG	240
CGTTCATAAC	AATCCC					256

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1328 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCCGGCCAC	CTCCGCTGCC	GCCAAAGCAA	CGGCACGGTT	CCACCGACGG	AAAAGTTTCT	60
GCGCCCCGAC	AGTCGGAGCA	TCATAAAAGA	CAGACCCGAC	CGCCAAGGCC	GCCACCGCCC	120
AAATTCGGGG	ATAGAACCGC	GGCCATCTC	TCGCAAAATA	TGCGGGACAT	GTACCTCGAT	180
ATGTGTACAT	CTTCGGGCCA	CAGGCCACGG	CCGCCAGCAC	CTCCGCGGCC	GAAAAAATGT	240
CAAACACACG	CCCCTCACCA	CGTTCATCAT	TGAAAGTCTC	TCCAGTCCAT	ATGTTGTCAG	300
GACGTGCTGT	CGTTCTCCGC	TTGCTGCGAA	GCCCGTTCTT	CCGAGTCGTG	TCGCTGCGTC	360
CAGCGTCGCG	CCCAAGATGG	GAATTTGGGT	CTTTTCACGC	GTAGCCTCCT	CCACCACGGC	420
TGCTGATCGC	CGTCACTAAG	GACCGACACG	GAGGATGACG	AGGAGCTTCT	CCCCGACTCC	480
GCGGTCCGCG	ACCGGCTACG	TAGCGCGTGT	CCCTGCCAGT	CTCCGCAGTT	ACACCACACG	540
TCGTGAGCAG	CGTGACCTG	CTGCCGCCAC	TGGGCCTCGG	CGTGCTCAGG	CCACCCGCCG	600
GAGCCCGGTC	TGAGCTCCGA	CGCAGGATGC	GCGTACTCAA	CGTGCGCCTT	CCAGTCCATA	660
CAGCAACACC	ATAGGTCGTG	CGAGTCGTG	GCTACCCGCC	GCCAGGCCAG	TTCCCGCATG	720
GGAAGGCTGG	ACACGCCGAC	CGAGAGGTCA	CCGAGCCCGG	ACGCCATCTC	TTCTTCCTCT	780
CCGTGCTGT	CATTAAGCAG	CCAGGTCACC	TCCTCCGCTC	CGCGTCCGCC	GGTCTCGACG	840

-continued

GACCGCGCCG	CCGTCTGGCAA	CACGGAAAAC	AGCACGCCAG	CCCGAGCCGC	TAAGGCCGCA	9 0 0
TGCCCCTGCC	GCCCAACTGA	ACACGCATAC	CCCGCTCAAC	TGCGTTTTGC	CACCCCTGTC	9 6 0
AGTGCTCTCG	CTCGAGCACC	ACCCCGCATC	TCCCAACCTT	TTTCCAATAA	ACGAAACCGA	1 0 2 0
CATGACACAC	GTAATGGGTA	CTCGTGGCTA	GATTTATTGA	AATAAACCGC	GATCCCGGGC	1 0 8 0
GTCTCAGCAC	ACGAAAAACC	GCATCCACAT	CATAGACAAG	TTACAGTCCA	CAGTCACATA	1 1 4 0
CACGATAAAC	AATACCAACA	GGGTAATGTT	TATGGAGTAA	AACACTATTG	TCCAGGCCAC	1 2 0 0
ATGCGTGTAT	GACTTCCGCA	CCATCCCGTA	CTGCATGTTC	CACATGTACG	CGCTAGACGT	1 2 6 0
GTAATCCACT	CGCAGTTCGG	GGACGCAACG	CAGCCAGATC	ACATCCCCTT	GCAGTACCAG	1 3 2 0
ACGCAGGG						1 3 2 8

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1528 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGGCCTGGGT	GTTAGAGTCC	GTACCTTGCA	GCCCAGGTAG	GTTTCAGGTA	CCAGCTGGTT	6 0
CGTACCTGTT	AAATAAATCG	CAGACGGGCG	CTCACCCCTA	CGGTCAGGAG	CACAAGAACA	1 2 0
ACCAGAGAGA	ACAGATATAC	GAGCAGGGTT	CTGAACAGCA	GACCCCAATT	GTCGTCTCTC	1 8 0
ATGCTTCGCT	GAAGGTACCA	GTTGATGGTC	TGAGAGCTAT	AGTCCATCCT	CACCTGAGGA	2 4 0
ACACACGCGG	CATATTTCTT	GGGGTCTCCC	CACCTCGTAG	ACAACGTGAT	GTCCACCATA	3 0 0
TCCACGGTGT	GCGTCACCGG	GTGCCACCGG	ATGTTCCACT	CGAAATAGGC	TCCGCGCTCA	3 6 0
TCATGGTGGT	ACTGCTCACC	GGACACCTGC	AGTCTGTCCA	TGTAAGATTG	AGAGACGATA	4 2 0
CCCACGTTCA	CAAAGTGTTT	CTCGGTGAAG	TTGCCCGACA	TCCTCCCCTT	GAAGTACAGC	4 8 0
ATGCCCATAT	GGAACCAGCA	TTGGTTCTCC	TCCACTCGAA	AGTGGGCCGA	TCTGATCTCC	5 4 0
GATACCACCA	CATCCAGGGG	CCGGGGCACC	GAGTCCGCGA	GTCTCAGGAA	CAAGACGGCC	6 0 0
AGGATCGCGA	GCACCAACAC	CGGCTTCATG	GCTCCGAAGG	TCCGCTGCTC	GGCTCCGCTC	6 6 0
ACCGCTCCGG	TCTGGCTGCA	GCAGTGCTTC	GCTGAGAAGT	AGCGTGTGGA	CTGAACGGTG	7 2 0
TTTTTGAATA	TATAGCGTTT	CTTGGTGACG	TTGTTTCCCC	TACGTAGTAG	GCAACTACGT	7 8 0
GCCAAAAGAG	GCGTTACGGT	ACTTTCGTA	CTGGGATTTT	CAAACCGGGA	CTTTCACAC	8 4 0
GGCGGTTTCA	ACACCGGGAC	TTTTCACACG	GTGATTTTCGG	CACCGGGACT	TTCCGCACGG	9 0 0
CGGTTTCGCC	ACCGCTGACG	TTCTCATCGC	CGCCCACGTC	AACGGTGGCG	ACACCGTACT	9 6 0
TTCCCATGCG	GTTTATAAAC	GTCAAGAGTC	ACGTCAGTCG	CCCACCCCA	TTACACGGCG	1 0 2 0
ATATCCCGAT	AGGGCATGAG	GGGACCCGGG	TGTCGCGACA	TGTCGACGAC	AGGTGCGGAT	1 0 8 0
TAGTGGTCGT	GTCGCGACAT	GGACGTGCAG	GGGGATGTCT	GTCGCGATAG	AGTTGATGTG	1 1 4 0
ACAGCCCGCT	ACACCTCTCT	GTCGCGACAT	GCATACACAA	CGGGCCGGCT	TGTCGGCGAT	1 2 0 0
TGTCGCGACA	TATCGTTATC	AGTTAGCGAC	CGGAGTTGTC	TATCGCGACA	TATCGTCGAC	1 2 6 0
TATCGCGACA	GAAAAAATAC	CGTTCGTAGA	GAATGCCGTG	TTGAAGGAAC	GCGCTTTTAT	1 3 2 0
TGAGACGATA	AAACAGCATC	AGGAGCCACA	ACGTCGAATC	CCACGTCCAG	TCGATTCGTA	1 3 8 0
TGTTATGCTG	CACAGCAATG	CTAGAATAAC	AACCAGCAGG	GTAATCCC GC	AACATAAATA	1 4 4 0

-continued

CAAAGTCACA	GCGAAGAATC	CGTGTCGTTC	TATCAAGCGA	AACGCGTTCC	AAACGGCCCC	1500
GTCACAGACG	CAGTTATTCA	TAAGCGTT				1528

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1151 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCATGCCGGG	AAAGCACATT	CCTTTTCAGT	AAACAACAAT	GACATCATAA	CAAATCATT	60
TATTCGCGAG	GTGGATAATA	ACCGCATATC	AGGAGGAGGG	ATCGGGTGAT	GACGCAGGCC	120
CCGCAGAACA	GTCCGAAATA	AATTTTTAGT	ATTGCCCCAT	AGTCGCCTAG	ATACCAGAGG	180
TACGTTAAGT	TCATCAAAAC	GCCCATCGGC	GTCCCAGGAAT	CGTATACCGG	GCACACGAAG	240
CGTTCATAAC	AATCCCGGGA	GGCGAGTGTT	AGGGTAGCAG	AGTAGTTTCG	GGGTCCGTTT	300
CCTTCCGGCG	ACGACAGTTC	CGTGGGCAGC	AGAATGTACA	GCGCCTCGGT	AGCTGTCGCG	360
GTGCCTTCCA	CGAGGATGGG	CTGCCGGTGC	CTTTCGTGAT	TTTCCCCGTC	GTGTAGCCAA	420
GCCGAGGCC	GCAAAGTCTT	AGGCGAGGGG	AATTGTCCAT	AGAGTTTCAC	CGCACCCCTC	480
AGTACATGGT	TCTGAATAAC	ACAGCCGCAC	GTGAAGTAGG	TAGGTTCTCT	CGTCTCCTCC	540
GTGGCTGCCG	CCACCACTCC	CAGCCACCAC	AACAGGCAGA	TCGCCAGAGG	GTTCCGGAGG	600
CTTCCCCGGC	GTAGCATGGT	TTTGGGTAA	AGCAAAAAGT	CTGGTGAGTC	GTTTCCGAGC	660
GACTCGAGAT	GCACTCCGCT	TCAGTCTATA	TATCACCACT	GGTCCGAAAA	CATCCAGGGA	720
AAATGTCGGT	GCAGCCAACC	TTTCACATAC	AGCCCCAAA	ACACTTGAAT	CACTGCCACC	780
ATCATCAGCG	TATACTGCGC	CGACTTAATC	GTGAGCGCGT	AGTACGCCAT	TAGACGGCGA	840
TCTTCGAACA	ATAGTCGTTC	GATGTCCTCT	AACGAGCTCC	ACAGGGGAAC	CCAAGGCACG	900
AGGCACCGGG	GTTTCGCACTC	TACATAATAA	GTTTGGCATT	GGTGGCAGGG	GGAAAAGTAG	960
AACAACACGA	GTTTTGTGCG	TTGGGGAACA	CGATAGTCCC	GGAGCCAGTA	GCGTTTTGCG	1020
ACGAGGCTTT	CGGAGACGTC	CTCCACCGGC	GTCGGCACTC	GATCCGCGTA	GCCCTCCAGC	1080
GTCTGGTAGT	ACACCCGGGG	TGTCGGCGTG	GGCACGGACA	GGTTCCCGCG	CAGGGTCCAC	1140
AGAGCCTCCA	G					1151

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1809 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCATTCTTTA	TGATAGGTGT	GACGATGTCT	TTACCATTCC	CACGGCTGCT	TTGCGTGATG	60
ATGACATTCA	TCATGTATTT	CCATTCACAC	ATACCTTTTG	TGCATACGGT	TTATATATGA	120
CCATCCACGC	TTATAACGAA	CCTAACAGTT	TATTAGCCCT	TGACAGGATA	GGTCAAAAGA	180
TTATATGTAG	GTTTTCCGGT	AAACCGAATT	GTGATATTTT	TCTGCAGGAA	ATAGAACAGC	240
CTGGTACCTA	TAAAACGGAC	AATGCAGTAC	TGTAGCAGCG	TAACCAAGTA	GGTCCACATG	300

-continued

AACACGTACA	AAATTATGGT	AAGCCATCGT	TTTTCATACC	ACAGCCTGTA	GCTGTCGTAC	3 6 0
ATGAATGAGG	ACGGTCGAGG	AACCCAGGGT	AGTTGTAATT	GGGGGCGACA	TTCGTACTGT	4 2 0
CCAGAAGACA	ATTGCACGGG	TTTCAGTGAG	ATGAGTACTT	TAGCGATGTC	GGCGGGGGCG	4 8 0
CTACGTTTTCA	CCGTGACGGT	GAGAACTTGA	CCGTGTTTTT	GTATTTTCATG	AGGCACGTTA	5 4 0
TACAAGCCAC	TGGTATCATG	AAGGATGACC	TCTGATGCGA	TGTGAGGATT	AAATTGTCCC	6 0 0
TCAAACCGCC	AAACGCTGGT	CATGTTTTCA	CCGTCAATTA	CGCAGCTGAC	GGTGTGAGAT	6 6 0
ACCACGATGT	TGGACTTAGG	TTTGGGGGCT	AATTGCCTTT	TTACAAATTC	CCTTCTGTAT	7 2 0
TGCAGGTCCT	GCTGCCACTG	CTTTTCCGTG	CGGAAAGTCG	CCATGTCTTC	CACACGTGTG	7 8 0
GCGACGATAG	ACGCCACCAA	GGTAGCTACC	AGAAGCAGCT	GGATCCGCAT	GGCATTACCG	8 4 0
TATGTCAATT	AGAAAGTTGA	GCGGACACGG	TTATCGTTCC	TGGCGGATAT	AAGTATATAA	9 0 0
ACGCGAGTTA	GCCTTTCCCG	TCCGTTTTGT	ACACCCGTTT	CCCACACAAA	TGACGAATAC	9 6 0
GACCTTTTTT	TTTATAAAAA	TAAACCACGT	GTATTATATA	AAAACATTTA	CATAGAAAAG	1 0 2 0
AGACACACGG	ATCAACATAA	GGACTTTTTCA	CACTTTTGGG	GTACACAGGC	GTGCCACCGC	1 0 8 0
AGATAGTAAG	CGCTGGATAC	ACGGTACACA	GTCCTGGCCA	GCACGTATCC	CAACAGCAGC	1 1 4 0
ACCATCGCCA	TACAGATGGC	GATCACGACC	CCGAGCTCTA	AGTGTCTGTA	TTCATAGTGT	1 2 0 0
AGTCGCCGCA	GGTTATCCAC	TGAATTCCCG	TAACTGAAAT	AACGTATATG	GTACCGAGGC	1 2 6 0
TGGCACCAACA	TGGGTTTGCA	TTTGGTGAC	GGCACCAAAT	GCAGAGTGAG	ATGGTCCAAG	1 3 2 0
TCCGTGGGCA	CCCCTGGCG	CAAACGGAAT	ACGGCTTCGG	TGGTCTCCAC	GAGGCACTCC	1 3 8 0
GGGGCGTGCA	GACGGCCCCA	CTTTCGTCCG	CGACGGCCCC	ACCAGCCGAC	CCGAGCCACT	1 4 4 0
ATCCCTTTCT	CGGGATAGAA	CGTACCCTGT	ACACGCCACA	CAGCGTCCAA	CACGCCGTCC	1 5 0 0
TTGACGACGC	AGCTGGCCTG	ATAGCTGGAC	ACGTTGTAA	GCGGCGGAAA	GCGAAACTGA	1 5 6 0
CGTGCCGGCG	GAGCCACATA	GTTTCGGTTC	CCGTGTTGTC	GCGGTTCTGTC	CTCCCTATAG	1 6 2 0
TAATAGTAGT	CGTCGTCCTC	ATAGGGGTTG	CCGGCGTGAG	CCAGCGTTAC	CCAACAGCAG	1 6 8 0
CCCAGGCCGA	CGAGGAGGCG	CAGCCACCGC	CTCATGGCGG	CTTCGCCAGT	CAATCGTCTT	1 7 4 0
TAGCCTCTTC	TTCCCGTGAG	GTCCTTCCGG	TGGCGCGGTG	CCGACCTCGG	ACCCAGGGAC	1 8 0 0
GTATCCACC						1 8 0 9

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1765 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCGCCGCCAG	CGGCGAGCGG	CACGGAGACG	GAGGCCGCCG	GCGGGGACGC	GCCGTGCGCG	6 0
ATAGCGGGAG	CCGTGGGGTC	CGCTGTACCT	GTGCCTCCGC	AGCCGTACGG	CGCCGCCGGC	1 2 0
GGGGGCGCGA	TTTGCGTGCC	TAACGCGGAC	GCGCACGCGG	TGGTCGGGGC	GGACGCGGCA	1 8 0
GCAGCAGCGG	CGCCGACGGT	GATGGTGGGT	TCGACAGCGA	TGGCGGGTCC	GGCGGCGTCG	2 4 0
GGGACCGTGC	CGCGCGCCAT	GCTGGTGGTG	CTGCTGGACG	AGCTGGGCGC	CGTGTTCTGGG	3 0 0
TACTGCCCGC	TGGACGGGCA	CGTGTACCCG	CTGGCGGCGG	AGCTGTCGCA	CTTTCTGCGC	3 6 0
GCGGGCGTGC	TGGGCGCGCT	GGCGCTGGGA	CGCGAGTCGG	CGCCCGCCGC	CGAGGCCGCG	4 2 0
CGGCGGCTGC	TGCCCGAGCT	GGACCGCGAG	CAGTGGGAGC	GGCCGCGCTG	GGACGCGCTG	4 8 0

-continued

CACCTGCACC	CGCGCGCCGC	GCTGTGGGCG	CGCGAGCCGC	ACGGGCAGTG	GGAGTTCATG	540
TTTCGCGAAC	AACGCGGTGA	CCCCATAAAT	GATCCCCTCG	CATTTTCGTCT	TTCGGACGCT	600
CGAACTCTCG	GTCTCGACCT	CACCACCGTC	ATGACAGAGC	GTCAAAGTCA	ATTGCCCGAA	660
AAGTATATCG	GTTTCTATCA	GATTAGGAAA	CCTCCTTGGC	TCATGGAACA	ACCTCCACCC	720
CCATCTCGCC	AAACCAAACC	GGACGCTGCA	ACGATGCCCC	CACCGCTCAG	TGCTCAGGCA	780
AGCGTCAGCT	ACGCGCTCCG	ATACGATGAC	GAGTCCTGGC	GCCCGCTCAG	CACAGTTGAC	840
GACCACAAAG	CCTGGTTGGA	TCTCGACGAA	TCACATTGGG	TCCTCGGGGA	CAGCCGACCC	900
GACGATATAA	AACAACGCAG	ACTGCTGAAG	GCCACTCAAC	GACGAGGCGC	CGAAATCGAC	960
AGACCCATGC	CTGTTCGTGCC	TGAAGAATGT	TACGACCAAC	GCTTCACTAC	CGAAGGCCAC	1020
CAGGTCATCC	CGTTGTGCGC	GTCCGAACCC	GAGGATGACG	ACGAAGATCC	TACCTACGAC	1080
GAATTGCCGT	CGCGCCCACC	CCAGAAACAT	AAGCCGCCAG	ACAAACCTCC	GCGCTTATGC	1140
AAAACGGGCC	CCGGCCCACC	TCCGCTGCCG	CCAAAGCAAC	GGCACGGTTC	CACCGACGGA	1200
AAAGTTTCTG	CGCCCCGACA	GTCGGAGCAT	CATAAAAGAC	AGACCCGACC	GCCAAGGCCG	1260
CCACCGCCCA	AATTCGGGGA	TAGAACCGCG	GCCCATCTCT	CGCAAAATAT	GCGGGACATG	1320
TACCTCGATA	TGTGTACATC	TTCGGGCCAC	AGGCCACGGC	CGCCAGCACC	TCCGCGGCCG	1380
AAAAAATGTC	AAACACACGC	CCCTCACCAC	GTTTCATCATT	GAAAGTCTCT	CCAGTCCATA	1440
TGTTGTCAGG	ACGTGCTGTC	GTTCTCCGCT	TGCTGCGAAG	CCCGTTCTTC	CGAGTCGTGT	1500
CGCTGCGTCC	AGCGTCGCGC	CCAAGATGGG	AATTTGGGTC	TTTTCACGCG	TAGCCTCCTC	1560
CACCACGGCT	GCTGATCGCC	GTCACTAAGG	ACCGACACGG	AGGATGACGA	GGAGCTTCTC	1620
CCCGACTCCG	CGGTCCGCGA	CCGGCTACGT	AGCGCGTGTC	CCTGCCAGTC	TCCGCAGTTA	1680
CACCACACGT	CGTGAGCAGC	GTGCACCTGC	TGCCGCCACT	GGGCCTCGGC	GTGCTCAGGC	1740
CACCCGCCGG	AGCCCGGTCT	GAGCT				1765

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1611 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GCCCCTGCCG	CCCAACTGAA	CACGCATACC	CCGCTCAACT	GCGTTTTGCC	ACCCCTGTCA	60
GTGCTCTCGC	TCGAGCACCA	CCCCGCATCT	CCCAACCTTT	TTCCAATAAA	CGAAACCGAC	120
ATGACACACG	TAATGGGTAC	TCGTGGCTAG	ATTTATTGAA	ATAAACCGCG	ATCCCGGGCG	180
TCTCAGCACA	CGAAAAACCG	CATCCACATC	ATAGACAAGT	TACAGTCCAC	AGTCACATAC	240
ACGATAAACA	ATACCAACAG	GGTAATGTTT	ATGGAGTAAA	ACACTATTGT	CCAGGCCACA	300
TGCGTGTATG	ACTTCCGCAC	CATCCCCTAC	TGCATGTTCC	ACATGTACGC	GCTAGACGTG	360
TAATCCACTC	GCAGTTCGGG	GACGCAACGC	AGCCAGATCA	CATCCCCTTG	CAGTACCAGA	420
CGCAGGGCTA	GCGTCTCGAA	GATCGGCATC	ACATCTAAGT	TCCGCACGTT	CCACTTTAAC	480
GACTCCCCGG	GAACGAACTC	CACGTCGTCG	GCGTGTACGT	ACAGGTTCTC	TCCCACGCCG	540
CCATAATCGG	CCTTCGGATC	GAAGACGAAC	CGACTCATGT	TGCCACGAT	GCTCCCCCGA	600
GCAAACAAC	TGCCGTTGTC	AATGTAGCAC	CGGTTGTCCT	CGATTTGAAA	CCAGGGATGC	660

-continued

TTGGCCGTGG	ACTTCCAGGG	CCGGAGCGCG	TCTTCCCCGG	CTTTAGTGAT	TCCATCGGGC	7 2 0
AGGCGGATCA	AGGGACCCAT	GGAGGTCCAA	AGACCCACCC	AGGCTTTCCA	GAGATTGTTC	7 8 0
ATGGTGAAAC	AGCGTGTGGA	CTGTACGCTC	TTTCCCAATT	TATATCCCAG	AGTAGTGACG	8 4 0
TGAGCCCAGC	CACCTCCCAG	ATTCCCTGACG	TTTTGGTTGT	CTTTCCTGCC	AATTCCTCCC	9 0 0
GTAAACTTAT	GATTATCCTA	GCCCATTTCC	GATAAAAATA	CACGGAGACA	GTAGATAGAG	9 6 0
TTACGAATAA	ACCGGTTTAT	TTATTCAAGT	GTCTCAGGAG	ATTATTGAAC	GAGCGTGGAT	1 0 2 0
ACCACGCCGT	CGTCAGTTCA	TGGTGGCATT	GAGCAGCCAT	AGCACCAGAG	TCCC GGCGCC	1 0 8 0
CGGTATCAGA	CACGCTGACC	TACCGGGCGC	CTTCGAGTCC	GTACCCCGCG	GCCTGGGTGT	1 1 4 0
TAGAGTCCGT	ACCTTGCAGC	CCAGGTAGGT	TTCAGGTACC	AGCTGGTTCG	TACCTGTAA	1 2 0 0
ATAAATCGCA	GACGGGCGCT	CACCCCTACG	GTCAGGAGCA	CAAGAACAAC	CAGAGAGAAC	1 2 6 0
AGATATACGA	GCAGGGTTCT	GAACAGCAGA	CCCCAATTGT	CGTCTCTCAT	GCTTCGCTGA	1 3 2 0
AGGTACCAGT	TGATGGTCTG	AGAGCTATAG	TCCATCCTCA	CCTGAGGAAC	ACACGCGGCA	1 3 8 0
TATTTCTTGG	GGTCTCCCCA	CCTCGTAGAC	AACGTGATGT	CCACCATATC	CACGGTGTGC	1 4 4 0
GTCACCGGGT	GCCCACCGAT	GTTCCACTCG	AAATAGGCTC	CGCGCTCATC	ATGGTGGTAC	1 5 0 0
TGCTCACCGG	ACACCTGCAG	TCTGTCCATG	TAAGATTGAG	AGACGATACC	CACGTT CACA	1 5 6 0
AAGTGTTTCT	CGGTGAAGTT	GCCCGACATC	CTCCCCTTGA	AGTACAGCAT	G	1 6 1 1

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1174 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCGCCGAAAT	CGACAGACCC	ATGCCTGTCTG	TGCCTGAAGA	ATGTTACGAC	CAACGCTTCA	6 0
CTACCGAAGG	CCACCAGGTC	ATCCCGTTGT	GCGCGTCCGA	ACCCGAGGAT	GACGACGAAG	1 2 0
ATCCTACCTA	CGACGAATTG	CCGTGCGGCC	CACCCAGAA	ACATAAGCCG	CCAGACAAAC	1 8 0
CTCCGCGCTT	ATGCAAAACG	GGCCCCGGCC	CACCTCCGCT	GCCGCCAAAG	CAACGGCACG	2 4 0
GTTCCACCGA	CGGAAAAGTT	TCTGCGCCCC	GACAGTCGGA	GCATCATAAA	AGACAGACCC	3 0 0
GACCGCCAAG	GCCGCCACCG	CCCAAATTCG	GGGATAGAAC	CGCGGCCCAT	CTCTCGCAA	3 6 0
ATATGCGGGA	CATGTACCTC	GATATGTGTA	CATCTTCGGG	CCACAGGCCA	CGGCCGCCAG	4 2 0
CACCTCCGCG	GCCGAAAAAA	TGTCAAACAC	ACGCCCTCA	CCACGTTTCA	CATTGAAAGT	4 8 0
CTCTCCAGTC	CATATGTTGT	CAGGACGTGC	TGTCGTTCTC	CGCTTGCTGC	GAAGCCCGTT	5 4 0
CTTCCGAGTC	GTGTGCTGTC	GTCCAGCGTC	GCGCCCAAGA	TGGGAATTTG	GGTCTTTTCA	6 0 0
CGCGTAGCCT	CCTCCACCAC	GGCTGCTGAT	CGCCGTCACT	AAGGACCGAC	ACGGAGGATG	6 6 0
ACGAGGAGCT	TCTCCCCGAC	TCCGCGGTCC	GCGACCGGCT	ACGTAGCGCG	TGTCCCTGCC	7 2 0
AGTCTCCGCA	GTTACACCAC	ACGTCGTGAG	CAGCGTGCAC	CTGCTGCCGC	CACTGGGCCT	7 8 0
CGGCGTGCTC	AGGCCACCCG	CCGGAGCCCG	GTCTGAGCTC	CGACGCAGGA	TGCGCGTACT	8 4 0
CAACGTGCGC	CTTCCAGTCC	ATACAGCAAC	ACCATAGGTC	GTGCGAGTCG	TCGGCTACCC	9 0 0
GCCGCCAGGC	CAGTTCCCGC	ATGGGAAGGC	TGGACACGCC	GACCGAGAGG	TCACCGAGCC	9 6 0
CGGACGCCAT	CTCTTCTTCC	TCTCCGTCGC	TGTCATTAAG	CAGCCAGGTC	ACCTCCTCCG	1 0 2 0
CTCCGCGTCC	GCCGGTCTCG	ACGGACCGCG	CCGCCGTCGG	CAACACGGAA	AACAGCACGC	1 0 8 0

-continued

CAGCCCGAGC CGCTAAGGCC GCATGCCCT GCGCCCAAC TGAACACGCA TACCCCGCTC 1140
 AACTGCGTTT TGCCACCCCT GTCAGTGCTC TCGC 1174

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 66 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCATGCCGGG AAAGCACATT CCTTTTCAGT AAACAACAAT GACATCATAA CAAATCATT 60
 TATTCG 66

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 645 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCACTGGTCC GAAAACATCC AGGGAAAATG TCGGTGCAGC CAACCTTTCA CATACAGCCC 60
 CCAAAACACT TGAATCACTG CCACCATCAT CAGCGTATAC TGCGCCGACT TAATCGTGAG 120
 CGCGTAGTAC GCCATTAGAC GCGGATCTTC GAACAATAGT CGTTCGATGT CCTCTAACGA 180
 GCTCCACAGG GGAACCCAAG GCACGAGGCA CCGGGGTTCG CACTCTACAT AATAAGTTTG 240
 GCATTGGTGG CAGGGGGAAA AGTAGAACAA CACGAGTTTT GTGCGTTGGG GAACACGATA 300
 GTCCCGGAGC CAGTAGCGTT TTGCGACGAG GCTTTCGGAG ACGTCCTCCA CCGGCGTCGG 360
 CACTCGATCC GCGTAGCCCT CCAGCGTCTG GTAGTACACC CGGGGTGTCG GCGTGGGCAC 420
 GGACAGGTTC CCGCGCAGGG TCCACAGAGC CTCCAGTCGA CCGCCCGATC GGAGCACGCA 480
 GCGCGCCTCG GAATACTCTA CTCGGTACTC CGAAACATCG GACAGAGGCG GTAACGGCTC 540
 CGTCTCCACC AAGGGCGGAG GTTCATCGAA AAGAGTCAAG GATAATTCAG GCATACTACC 600
 CGCGACCGGG GCCCAGAGGG CTAGAATAAG CATTACAAGG TTCAT 645

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 709 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: DNA (genomic)

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CTCGAGATGC ACTCCGCTTC AGTCTATATA TCACCACTGG TCCGAAAACA TCCAGGGAAA 60
 ATGTCGGTGC AGCCAACCTT TCACATACAG CCCCCAAAAC ACTTGAATCA CTGCCACCAT 120
 CATCAGCGTA TACTGCGCCG ACTTAATCGT GAGCGCGTAG TACGCCATTA GACGGCGATC 180
 TTCGAACAAT AGTCGTTTGA TGTCCTCTAA CGAGCTCCAC AGGGGAACCC AAGGCACGAG 240
 GCACCGGGGT TCGCACTCTA CATAATAAGT TTGGCATTGG TGGCAGGGGG AAAAGTAGAA 300

-continued

CAACACGAGT	TTTGTGCGTT	GGGGAACACG	ATAGTCCCGG	AGCCAGTAGC	GTTTTGCGAC	3 6 0
GAGGCTTTTCG	GAGACGTCCT	CCACCGGCGT	CGGCACTCGA	TCCGCGTAGC	CCTCCAGCGT	4 2 0
CTGGTAGTAC	ACCCGGGGTG	TCGGCGTGGG	CACGGACAGG	TTCCC GCGCA	GGGTCCACAG	4 8 0
AGCCTCCAGT	CGACCGCCCG	ATCGGAGCAC	GCAGCGCGCC	TCGGAATACT	CTACTCGGTA	5 4 0
CTCCGAAACA	TCGGACAGAG	GCGGTAACGG	CTCCGTCTCC	ACCAAGGGCG	GAGGTTTCATC	6 0 0
GAAAAGAGTC	AAGGATAATT	CAGGCATACT	ACCCGCGACC	GGGGCCAGCA	GGGCTAGAAT	6 6 0
AAGCATTACA	AGGTTTCATTC	TGTCTTACAA	GGGAAGGCTG	TTACCCTGT		7 0 9

What is claimed is:

1. A recombinant human cytomegalovirus comprising a genome from which the gene sequence encoding open reading frames IRS-1 to US11 has been deleted.

2. A recombinant human cytomegalovirus comprising a genome having a deletion of the gene sequence encoding open reading frames US2 to US11.

15

3. A recombinant human cytomegalovirus comprising a genome having a deletion of the gene sequences encoding open reading frames US2 to US5 and US10 to US11.

4. A recombinant human cytomegalovirus comprising a genome having a deletion of the gene sequences encoding open reading frames US2 to US5 and US11.

20

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,846,806
DATED : December 8, 1998
INVENTOR(S) : Jones et al.

Page 1 of 1

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 1,
Line 4, after the title, insert

-- STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER
FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

The present invention was made with Government support under Grant No. CA41451 awarded by the National Institutes of Health. The Government has certain rights in the invention. --.

Signed and Sealed this

Twenty-ninth Day of July, 2003

A handwritten signature in black ink, appearing to read "James E. Rogan", with a horizontal line drawn underneath it.

JAMES E. ROGAN
Director of the United States Patent and Trademark Office